EDK | CDIP | CDPE | CDEP |
Schweizerische Konferenz der kantonalen Erziehungsdirektoren
Conférence suisse des directeurs cantonaux de l'instruction publique
Conferenza svizzera dei direttori cantonali della pubblica educazione
Conferenza svizra dals directurs chantunals da l'educaziun publica

GESAMTSCHAU ZUR ÜBERPRÜFUNG DES ERREICHENS DER GRUNDKOMPETENZEN (ÜGK)

Bericht des Koordinationsstabs für die Umsetzung der Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS)

23. November 2022

222.0-4.5 pl

Generalsekretariat | Secrétariat général
Haus der Kantone, Speichergasse 6, Postfach, CH-3001 Bern | T: +41 31 309 51 11, F: +41 31 309 51 50, www.edk.ch, edk@edk.ch

IDES Informationszentrum | Centre d'information | T: +41 31 309 51 00, F: +41 31 309 51 10, ides@edk.ch

Inhalt

Einleitung 2

1 Bestandesaufnahme 3
1.1 Grundlagen und Vorarbeiten zur ÜGK 3
1.2 Umsetzung der ÜGK 9
1.3 Durchgeführte ÜGK-Erhebungen 13
1.4 Beschlossene zukünftige ÜGK-Erhebungen 20

2 Analyse und Diskussion 24
2.1 Ansatz, Ziele und Konzepte der ÜGK 24
2.2 Erhebungsdesign der ÜGK 28
2.3 Weitere Aspekte der Umsetzung der ÜGK 31
2.4 Governance der ÜGK 31

3 Vorgehen im Hinblick auf die Weiterentwicklung der ÜGK 35
3.1 Neues Erhebungsdesign 35
3.2 Anpassungen im Bereich der Governance 43
3.3 Kostenfolgen 46

Quellenverzeichnis 47

Anhang 1: Weitere Aspekte der Umsetzung der ÜGK 51

Anhang 2: Gutachten zur ÜGK 59

Einleitung

Am 25. Oktober 2019 erteilte die Plenarversammlung der EDK dem Koordinationsstab für die Umsetzung der Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS) den Auftrag, «im Hinblick auf den Entscheid der Plenarversammlung zur Weiterführung der Überprüfung des Erreichens der Grundkompetenzen nach 2022» eine «Gesamtschau» zu erstellen. Es war vorgesehen, bis ins Jahr 2022 im Rahmen der Überprüfung des Erreichens der Grundkompetenzen (ÜGK) vier Haupterhebungen durchzuführen (2016: 11. Schuljahr, Mathematik; 2017: 8. Schuljahr, Schulsprache und erste Fremdsprache; 2020: 11. Schuljahr, Schulsprache, erste und zweite Fremdsprache; 2022: 4. Schuljahr, Schulsprache und Mathematik). Wegen der Corona-Pandemie und wegen der Terminkoordination mit der PISA-Studie mussten die dritte und vierte Erhebung auf die Jahre 2023 bzw. 2024 verschoben werden.

Die vorliegende Gesamtschau gliedert sich in die Hauptteile «Bestandesaufnahme», «Analyse und Diskussion» sowie «Vorgehen im Hinblick auf die Weiterentwicklung der ÜGK». Die Bestandesaufnahme vermittelt einen Überblick über die Grundlagen und die Umsetzung der ÜGK bis hin zu der geplanten Erhebung im 4. Schuljahr. Der zweite Hauptteil ist eine kritische Auseinandersetzung mit verschiedenen Aspekten der ÜGK, die im Hinblick auf deren Weiterführung einer Klärung bzw. Konsolidierung bedürfen. Als Grundlage dafür hat der Kosta HarmoS von externen Expertinnen und Experten aus dem In- und Ausland ein umfassendes Gutachten eingeholt und zudem weitere Expertinnen und Experten Stellung nehmen lassen, die in unterschiedlichem Umfang in die ÜGK involviert sind oder waren. Im dritten Hauptteil werden die Eckwerte für die Weiterführung der ÜGK vorgestellt, die der Kosta HarmoS den Gremien der EDK zur Beurteilung und Beschlussfassung unterbreitet. Dabei handelt es sich im Wesentlichen um die Wahl des zukünftigen, langfristig angelegten Erhebungsdesigns sowie um Änderungen im Bereich der Governance der ÜGK.

Die Gesamtschau richtet sich an die Mitglieder der EDK. Sie soll ihnen als Grundlage und Argumentarium für Entscheidungen zur Weiterführung der ÜGK dienen.

1 Bestandesaufnahme

Der Teil «**Bestandesaufnahme**» stellt Folgendes dar: die konzeptionellen, politischen und rechtlichen Grundlagen der ÜGK; die Strukturen und Akteure zu deren Umsetzung; die vergangenen und in Entwicklung befindlichen ÜGK-Erhebungen; konzeptionelle Anpassungen aufgrund der Erfahrungen; die Berichterstattung zu den Ergebnissen der Erhebungen und die Nutzung der Daten der Erhebungen; den finanziellen Aufwand.

1.1 Grundlagen und Vorarbeiten zur ÜGK

1.1.1 Stärkung der Bildungskooperation in der Schweiz durch einheitliche Kompetenzniveaus

Nach ersten Diskussionen im Jahr 2000 verabschiedet die Plenarversammlung der EDK am 7. Juni 2001 unter ihrem neuen Generalsekretär Hans Ambühl (seit Januar 2000 im Amt) ein Leitbild («Leitlinien», EDK, 2001a) und erstmals ein Tätigkeitsprogramm (EDK, 2001b). In Zentrum des Leitbildes steht die Stärkung der Bildungskooperation in der Schweiz mit dem Ziel, die Wettbewerbsfähigkeit zu erhalten, den Austausch zwischen den Kulturen zu fördern und die Mobilität im Land zu erleichtern. Zur Stärkung der Kooperation und Koordination auf interkantonaler Ebene werden eine Erweiterung des Schulkonkordats von 1970 (EDK, 1970) und eine Reihe von koordinierenden Massnahmen ins Auge gefasst. Unter anderem sollen die Übergänge zu den Sekundarstufen I und II verbindlicher harmonisiert werden, indem für das Ende des 6. und des 9. Schuljahres (8. und 11. Schuljahr nach HarmoS-Zählung) landesweit einheitliche Kompetenzniveaus, sog. Treffpunkte, festgelegt werden. Weiter sollen zwecks Qualitätssicherung kontinuierliche Kompetenzmessungen eingeführt werden, insbesondere am Ende der obligatorischen Schulzeit. Gemeinsame Bildungsziele und die stetige Überprüfung der Zielerreichung gelten ab den Nullerjahren als zentrales Element der Harmonisierung des schweizerischen Bildungssystems. Verbindlich festgelegte Kompetenzniveaus im Sinne von Standards sollen, sobald sie vorliegen, dem Schulkonkordat von 1970 als Ergänzung hinzugefügt werden.

Die Einführung von Bildungszielen bzw. «Standards» und von Systemen zu deren Überprüfung ist zu dieser Zeit auch in anderen Ländern im Gange. Auslöser dafür war insbesondere die PISA-Studie, die im Jahr 2000 erstmals durchgeführt wurde und vielerorts mit Ergebnissen überraschte, die schlechter ausfielen als erwartet. Auch in der Schweiz empfiehlt die Steuerungsgruppe, die sich mit den Lehren aus PISA 2000 befasst, fachliche Leistungen und fächerübergreifende Kompetenzen in mehrjährigen Intervallen aufgrund von Bildungsstandards zu evaluieren (Buschor et al., 2003:31).

1.1.2 Das Projekt HarmoS

Am 6. Juni 2002 stimmt die EDK-Plenarversammlung der Organisation und Finanzierung des Projekts HarmoS zu. Eine erste Phase ist auf die Jahre 2003-2006 angelegt (vgl. EDK, 2004b:1). Der Auftrag umfasst vorerst drei Etappen:

1. Entwicklung von Kompetenzmodellen für vier Fächer (Kompetenzniveaus, Niveaubeschreibungen);
2. Empirische Überprüfung der Erfüllung der Anforderungen anhand von Stichproben von Schülerinnen und Schülern;
3. Festlegung von verbindlichen Standards durch die Bildungspolitik.

Es ist vorgesehen, eine zweite Phase daran anzuschliessen, um die Standards in Lehrplänen, Lehrmitteln sowie auf unterschiedlichen Ebenen des Bildungssystems in Beurteilungs- und Testinstrumenten umzusetzen.

Anfang 2003 nimmt das Projekt unter der Ägide eines Koordinationsausschusses bestehend aus Vertretern des Generalsekretariat der EDK (GS EDK) und den vier EDK-Regionalsekretären seine Arbeit auf. Ein Beirat, bestehend aus Vertreterinnen und Vertretern der vier EDK-Regionen, der Lehrerinnen und Lehrer sowie der Wissenschaft steht ihm beratend zur Seite.

Den Arbeiten in den genannten Etappen gehen grundlegende Diskussionen zur Art und Weise voraus, wie die Leistungserwartungen an die Schülerinnen und Schüler, d.h. die Bildungsstandards, beschrieben werden sollen. Die Klärung dieser Frage ist für alle drei Etappen von grundlegender Bedeutung. Schliesslich orientiert man sich an der in Deutschland im Jahr 2003 erstmals erschienenen Expertise «Zur Entwicklung nationaler Bildungsstandards», die unter der Leitung des Bildungsforschers E. Klieme entstanden ist, (vgl. EDK, 2004b, S. 7-11; Klieme et al., 2004) und konzipiert die schweizerischen Bildungsstandards als ergebnis- oder outputorientierte Leistungsstandards (auch «Performance-» oder «Outcome-Standards»), die auf die Kernbereiche der Fächer fokussieren, die auch in den kantonalen und regionalen Lehrplänen wiederzufinden sind. Für diese Kernbereiche sollen im HarmoS-Projekt Kompetenzmodelle entwickelt werden, die über die Stufen hinweg Entwicklungsverläufe aufzeigen. Bestimmte Punkte auf den so definierten Skalen sollen schliesslich zu Mindeststandards erklärt werden. Mindeststandards beschreiben Mindestkompetenzen, die von (praktisch) allen Schülerinnen und Schülern zu gegebenen Zeitpunkten verbindlich erwartet werden können. Die Expertise von Klieme et al. schlägt neben solchen verbindlichen Mindeststandards eine weitere Differenzierung der Anforderungen in Form von Kompetenzstufen vor. Dadurch sollen Lernentwicklungen aufgezeigt sowie Abstufungen und Profilbildungen ermöglicht werden. In der Schweiz verzichtet man auf nationaler Ebene auf Stufenbeschreibungen (z.B. für die leistungsdifferenzierten Modelle auf der Sekundarstufe I) und überlässt diese Möglichkeit den Kantonen. Verbindliche nationale Kompetenzmodelle und Mindest-Leistungsstandards (die späteren Grundkompetenzen) gewährleisten eine Harmonisierung der betreffenden Fächer bezüglich ihrer Kernbereiche, lassen aber wesentliche Aspekte ungeregelt, was auf den untergeordneten Ebenen bis hin zum Unterricht Freiräume gewährt. Bei der Entwicklung der sprachregionalen Lehrpläne bestand der Auftrag, die nationalen Bildungsziele bei der Beschreibung der Grundansprüche des Lehrplans, wenn möglich, wörtlich zu übernehmen. Lerninhalte, die über den Kernbereich hinausgehen, Unterrichts- und Lernprozesse sowie Lernbedingungen können aus nationaler Sicht auf der Lehrplan-, Lehrmittel- und Unterrichtsebene frei gestaltet werden, solange das Erreichen der Mindeststandards gewährleistet ist. Leistungsstandards lassen nicht bloss Freiräume hinsichtlich der Lehrplan- und Unterrichtsgestaltung zu, sondern eignen sich auch für die Ableitung von Leistungszielen und als Grundlage für die Entwicklung von Tests, beispielsweise von Tests, welche das Erreichen von Mindeststandards überprüfen.

Das Projekt HarmoS beschränkt sich von Anfang an auf die Entwicklung von Standards zu vier Fachbereichen, die als grundlegend erachtet werden und deshalb auch im Hinblick auf das Ziel, die Mobilität zwischen den Kantonen zu erleichtern, besonders relevant erscheinen: Schulsprache, Fremdsprachen, Mathematik und Naturwissenschaften. Was die Sprachen angeht, sind bereits seit den 1990er Jahren Arbeiten zur Koordination des Sprachenunterrichts im Gange. Am 25. März 2004 verabschiedet die EDK-Plenarversammlung schliesslich ihre Sprachenstrategie (EDK, 2004a), die u.a. die Festlegung von überprüfbaren und verbindlich zu erreichenden Kompetenzniveaus in der Schulsprache und in den Fremdsprachen als eines der Koordinationsinstrumente vorsieht. Bezüglich der Beschreibung und validen Überprüfung von Kompetenzen in den Fachbereichen Schulsprache, Mathematik und Naturwissenschaften liegen zudem aus der PISA-Studie nutzbare Erfahrungen vor. Der Einbezug weiterer Fachbereiche – zur Diskussion stehen in einer frühen Phase etwa ein musisches Fach oder Sport – wird vorerst verschoben, insbesondere auch um das Projekt nicht zu überladen (vgl. auch Abschnitt 1.1.4: Umsetzungsbeschluss zum HarmoS-Konkordat).

Für die Umsetzung der ersten und zweiten Etappe des HarmoS-Projekts arbeitet die EDK mit Fachdidaktikkonsortien der vier ausgewählten Fachbereiche zusammen, in der zweiten Etappe zudem mit einer Methodologiegruppe, die ihre Kompetenzen aus der Durchführung der PISA-Studien einbringt. Die überregional zusammengesetzten Konsortien erarbeiten Kompetenzmodelle zu dem, was übeeinstimmend als Kernbereiche ihrer Fächer wahrgenommen wird, und setzen Teile dieser Modelle in Aufgaben um. Im Frühjahr 2007 wird in der Deutschschweiz und in der Westschweiz in regional repräsentativen Stichproben von Schülerinnen und Schülern am Ende der Primarstufe und am Ende der Sekundarstufe I eine Leistungsmessung durchgeführt. Das Ziel besteht in der Validierung der Kompetenzmodelle und der Aufgaben. Aus zeitlichen und finanziellen Gründen muss auf gründliche Analysen der Ergebnisse sowie

auf einen Schwellenwertsetzungsprozess nach wissenschaftlichen Standards verzichtet werden. Weil in der Validierungsstudie nicht spezifisch auf die Mindestkompetenzen fokussiert wird, ist es anschliessend auch nicht möglich, aus Testaufgaben und Testergebnissen im erhofften Umfang Informationen darüber zu gewinnen, was mindestkompetente Schülerinnen und Schüler tatsächlich können (vgl. Schlussbericht der Methodologiegruppe: Ramseier et al., 2008). Die Konsortien schlagen trotzdem nach einem Konsolidierungsprozess Mindeststandards für die 2. (ausser bei den Fremdsprachen), 6. und 9. Klasse (nach alter Zählweise) vor (vgl. dazu die wissenschaftlichen Berichte der Konsortien: Konsortium HarmoS Fremdsprachen et al., 2009; Konsortium HarmoS Naturwissenschaften+, 2010; Konsortium HarmoS Schulsprache, 2010; Konsortium Mathematik et al., 2009).

Die dritte Etappe des HarmoS-Projekts, die Festlegung von verbindlichen Standards durch die Bildungspolitik, kann erst zu einem späteren Zeitpunkt abgeschlossen werden. Vorerst werden im Jahr 2010 die Leistungsstandards, welche die Konsortien erarbeitet haben, in eine Anhörung geschickt (vgl. Art. 7 des HarmoS-Konkordats). Es folgt eine Überarbeitung in Zusammenarbeit mit den Fachdidaktikkonsortien, und schliesslich werden die nationalen Bildungsstandards unter der Bezeichnung «Grundkompetenzen» am 16. Juni 2011 durch die Plenarversammlung freigegeben. Für jeden der vier Fachbereiche wird eine Broschüre publiziert, die im Wesentlichen die Grundkompetenzen beschreibt, die am Ende der Bildungsstufen zu erreichen sind (EDK, 2011a, 2011d, 2011b, 2011c). Die freigegebenen Bildungsziele richten sich in erster Linie an Fachleute, die Lehrpläne, Lehrmittel oder Beurteilungsinstrumente erarbeiten. Sie sollen als Zielvorgaben in den bestehenden *Plan d'études romand* sowie die geplanten Lehrpläne *Lehrplan 21* der Deutschschweiz[1] und *Piano di studio* des Kantons Tessin einfliessen, was dann auch der Fall ist.

Die Grundkompetenzen beschreiben die Kompetenzen, die am Ende der drei Bildungsstufen der obligatorischen Schule (4., 8. und 11. Schuljahr, Fremdsprachen 8. und 11. Schuljahr) von jeweils (praktisch) allen Schülerinnen und Schülern zu erreichen sind. Zu jedem Fachbereich wird in den Broschüren zu den Bildungszielen ein Kompetenzmodell beschrieben, das der Gliederung des Fachbereichs und damit u.a. auch der Planung von Tests dient. Die vertikale Dimension des Kompetenzzuwachs wird dabei erst über die Schulstufen hinweg sichtbar. Pro Stufe wird bloss beschrieben, was es heisst, über die Grundkompetenzen zu verfügen. Es wird nicht weiter differenziert[2]. Konsequenterweise sollen die standardsbasierten Tests dann auch (nur) das Erreichen der Schwelle «Grundkompetenzen erreicht» überprüfen.

1.1.3 Bildungsartikel der Bundesverfassung

Parallel zu den Arbeiten an einer Ergänzung zum Schulkonkordat auf interkantonaler Ebene sind von Anfang an auch auf Bundesebene Bestrebungen zu einer Harmonisierung der Schulsysteme in der Schweiz im Gange. Sie führen schliesslich zu einer Verfassungsänderung. Am 21. Mai 2006 nimmt das Schweizer Stimmvolk die revidierten Bildungsartikel in der Bundesverfassung an. Seither sind die Kantone unter anderem dazu verpflichtet, wichtige Eckwerte der obligatorischen Schule national zu harmonisieren. Gegenstand der Harmonisierung sind das Schuleintrittsalter, die Schulpflicht, die Dauer und die Ziele der Bildungsstufen, die Übergänge von einer Stufe zur anderen sowie die Anerkennung von Abschlüssen (Artikel 62 Absatz 4 Bundesverfassung). Diese Bestimmung enthält weiter eine subsidiäre Bundeskompetenz zum Erlass der notwendigen Vorschriften, wenn die Harmonisierung auf dem Weg der interkantonalen Koordination nicht gelingen sollte. Der Auftrag, die Ziele der Bildungsstufen zu harmonisieren, ist für die Entwicklung und Umsetzung der Bildungsziele und der ÜGK von zentraler Bedeutung.

---

[1] Für die romanischsprachige Schweiz wird eine spezifische Fassung des Lehrplan 21 erarbeitet.

[2] Ursprünglich bestand die Absicht, neben den Grundkompetenzen auch höhere Kompetenzstufen zu beschreiben. Die Fachkonsortien haben teilweise entsprechende Vorschläge vorgelegt. Hauptsächlich aus Ressourcengründen konnten diese Arbeiten aber nicht zu Ende geführt werden. Es gelang auch nicht, die Beschreibungen der Bildungsziele für die vier Fachbereiche gleichartig zu gestalten.

1.1.4 HarmoS-Konkordat

Rund ein Jahr später, am 14. Juni 2007, verabschiedet die Plenarversammlung der EDK die Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat; EDK, 2007a)[3], die in Artikel 7 die Festlegung von nationalen Bildungsstandards vorsieht.

HarmoS-Konkordat

Art. 7 Bildungsstandards

[1] Zur gesamtschweizerischen Harmonisierung der Unterrichtsziele werden nationale Bildungsstandards festgelegt.

[2] Unterschieden wird zwischen folgenden zwei Arten von Bildungsstandards:

a. Leistungsstandards, die pro Fachbereich auf einem Referenzrahmen mit Kompetenzniveaus basieren;
b. Standards, welche Bildungsinhalte oder Bedingungen für die Umsetzung im Unterricht umschreiben.

[3] Die nationalen Bildungsstandards werden unter der Verantwortung der EDK wissenschaftlich entwickelt und validiert. Sie unterliegen einer Vernehmlassung gemäss Artikel 3 des Konkordats über die Schulkoordination vom 29. Oktober 1970.

[4] Sie werden von der Plenarversammlung der EDK mit einer Mehrheit von zwei Dritteln ihrer Mitglieder verabschiedet, von denen mindestens 3 einen nicht mehrheitlich deutschsprachigen Kanton vertreten. Die Revision erfolgt durch die Vereinbarungskantone in einem analogen Verfahren.

Art. 8 Lehrpläne, Lehrmittel und Evaluationsinstrumente

[1] Die Harmonisierung der Lehrpläne und die Koordination der Lehrmittel erfolgen auf sprachregionaler Ebene.

[2] Lehrpläne, Lehrmittel und Evaluationsinstrumente sowie Bildungsstandards werden aufeinander abgestimmt.

[3] Die Kantone arbeiten im Rahmen des Vollzugs dieser Vereinbarung auf sprachregionaler Ebene zusammen. Sie können die hierfür erforderlichen Einrichtungen schaffen.

[4] Die EDK und die Sprachregionen verständigen sich von Fall zu Fall über die Entwicklung von Referenztests auf Basis der Bildungsstandards.

Art. 10 Bildungsmonitoring

[1] In Anwendung von Artikel 4 des Konkordats über die Schulkoordination vom 29. Oktober 1970 beteiligen sich die Vereinbarungskantone zusammen mit dem Bund an einem systematischen und kontinuierlichen, wissenschaftlich gestützten Monitoring über das gesamte schweizerische Bildungssystem.

[2] Die Entwicklungen und Leistungen der obligatorischen Schule werden regelmässig im Rahmen dieses Bildungsmonitorings evaluiert. Ein Teil davon ist die Überprüfung der Erreichung der nationalen Bildungsstandards namentlich durch Referenztests im Sinne von Artikel 8 Absatz 4.

In Art. 7 Abs. 2 des Konkordats ist von zwei Typen von Standards die Rede: einerseits von Leistungsstandards wie den nationalen Bildungszielen, andererseits von Standards, die sich auf Lerninhalte und Unterrichtsgestaltung beziehen. Standards vom zweiten Typ sollen in Fachbereichen wie dem bildnerischen Gestalten oder Sport umgesetzt werden, für die Leistungsstandards kaum Sinn machen.

Art. 7 Abs. 4 des Konkordats sieht ein Verfahren für die Revision der Bildungsziele vor. In Anbetracht der ständigen Entwicklung des Bildungssystems ist dies grundsätzlich notwendig, es kann aber auch genutzt werden, wenn sich zum Beispiel die freigegebenen Bildungsziele bei der Überprüfung der

[3] Für Hintergründe und Entstehungsgeschichte des HarmoS-.Konkordats siehe den Kommentar des damaligen Generalsekretärs der EDK (Ambühl, 2011).

Grundkompetenzen nicht vollends bewährt haben. In diesem Zusammenhang ist daran zu erinnern, dass die Bildungsziele bei den Fremdsprachen zu einem Zeitpunkt definiert wurden, als der frühere Beginn des Fremdsprachenunterrichts noch nicht umgesetzt war und die Folgen nur schwer abgeschätzt werden konnten.

Art. 8 schafft die Grundlage für die anschliessend erfolgte bzw. zu Ende geführte (Westschweiz) Entwicklung gemeinsamer sprachregionaler Lehrpläne und überträgt den regionalen bzw. subregionalen EDK-Konferenzen die Verantwortung dafür sowie auch für die Lehrmittelkoordination. Die gemeinsamen Bildungsstandards werden in die Lehrpläne integriert.

Art. 10 schafft in Verbindung mit Art. 8 Abs. 4 des Konkordats die Grundlage für die Entwicklung und Durchführung der ÜGK. Sie soll Teil des systematischen und kontinuierlichen, wissenschaftlich gestützten Monitorings im Bereich der obligatorischen Schule werden.

Das HarmoS-Konkordat tritt am 1. August 2009 in Kraft. Nachdem das Konkordat von einer Mehrheit der Kantone, in denen gegen 80% der schweizerischen Wohnbevölkerung leben, ratifiziert ist, richtet sich die verfassungsmässige Harmonisierung der obligatorischen Schule hinsichtlich der strukturellen Eckwerte und der nationalen Bildungsziele grundsätzlich nach den im Konkordat vorgesehenen Lösungen. Nach aktuellem Stand sind 15 Kantone dem Konkordat beigetreten (vgl. EDK, 2021a).

Der *Umsetzungsbeschluss* der Plenarsammlung vom 25./26. Oktober 2007 zum HarmoS-Konkordat (EDK, 2007b) bestimmt die Bildungsziele und die Referenztests näher.

- Die nationalen Bildungsziele legen demnach Standards für die Fachbereiche Schulsprache, Fremdsprachen, Mathematik und Naturwissenschaften fest sowie später, und basierend auf den ersten Erfahrungen, Standards für weitere Fächer: prioritär ICT, Bewegungserziehung und -förderung, Musik, Bildnerisches Gestalten (Ziff. 5.1, 3.1).
- Die Referenztests (Ziff. 5.3 – 5.7) werden für zwei Funktionen geschaffen:
  1) Systemevaluation auf nationaler und regionaler bzw. kantonaler Ebene im Rahmen des Bildungsmonitorings». Dabei liegt die Erarbeitung der Instrumente in der Verantwortung der EDK (Ziff. 5.3.1).
  2) Individuelle Standortbestimmung zur Förderplanung, vor allem im Hinblick auf den Übertritt in die Sekundarstufe II. Die Erarbeitung dieser Instrumente wird durch die EDK koordiniert, liegt aber in der Verantwortung der Regionen.

- Die Systemevaluation durch die ÜGK und die PISA-Studie sollen künftig komplementär angelegt sein, d.h. PISA soll nach der Umsetzung der ÜGK nur noch dem internationalen, nicht mehr dem interkantonalen Vergleich dienen[4]. Im Beschluss selbst wird nicht explizit darauf hingewiesen, dass sich bei interkantonalen Vergleichen durch die ÜGK auch das Vergleichskriterium ändert: Während in PISA die Schülerinnen und Schüler auf Kompetenzskalen verortet werden, wird im Rahmen der ÜGK der Anteil der Schülerinnen und Schüler verglichen, welche die Grundkompetenzen erreichen bzw. nicht erreichen.
- Die Referenztests dürfen nicht für Schulrankings und die Leistungsbeurteilung von Lehrpersonen verwendet werden.

Exkurs zu den individuellen Standortbestimmungen: Die im Umsetzungsbeschluss erwähnten Standortbestimmungen zur individuellen Förderplanung, deren Erarbeitung in der Verantwortung der Regionen liegt, sind in den Regionen und Kantonen bisher in unterschiedlicher Form und nicht flächendeckend umgesetzt worden. Zudem steht bei kantonalen oder vom Kanton geförderten Tests nicht immer die individuelle Förderplanung im Vordergrund, sondern beispielsweise das Monitoring des Schulsystems. Zur Zeit der Einrichtung der Aufgabendatenbank (ADB) der EDK ging die EDK noch davon

[4] Sobald die Teilnahme an der PISA nur noch dem internationalen Vergleich dient, beteiligt sich die Schweiz auch nur noch mit einer nationalen Stichprobe der 15-Jährigen an PISA, wie das international üblich ist, und nicht mehr mit einer erweiterten Stichprobe, die gleichzeitig auch in jedem Kanton für das 11. Schuljahr repräsentativ ist.

aus, dass die heute *Checks* genannten Tests des Bildungsraums Nordwestschweiz, das Testangebot *Stellwerk* des Lehrmittelverlags St. Gallen und die ADB kompatible Formate verwenden würden, um einen Austausch von Aufgaben zu ermöglichen. Diese Idee hat sich aber nicht durchgesetzt. Die ADB konnte bei den Standortbestimmungen bis heute keine Rolle spielen. Die D-EDK versuchte um das Jahr 2012, eine «Koordination Leistungsmessung Deutschschweiz» umzusetzen. Dies gelang aber nicht. Die Koordination erfolgte kleinräumiger. Der Bildungsraum Nordwestschweiz (Kantone Aargau, Basel-Landschaft, Basel-Stadt und Solothurn) setzt mit den *Checks* ein umfassendes Testsystem für wiederholte individuelle Standortbestimmungen ein. Auch der Kanton Zug nutzt die Checks punktuell. Aufgabensammlungen unter der Bezeichnung *Mindsteps* dienen als nicht-obligatorische Ergänzung zu den *Checks* für die formative Nutzung durch die Schulen. Analog zum Angebot aus *Checks* und *Mindsteps* haben die Kantone St. Gallen und Zürich *Stellwerk* zu den Lern- und Testsystemen *Lernlupe* (Primarstufe) und *Lernpass Plus* (Sekundarstufe I) ausgebaut. Diese beiden Kantone, sowie zahlreiche weitere Deutschschweizer Kantone sowie das Fürstentum Liechtenstein[5] nutzen *Stellwerk* (jetzt als Teil von *Lernpass Plus*) für Standortbestimmungen, oft im Zusammenhang mit der Berufswahl. Der Einsatz von *Stellwerk* erfolgt je nach Kanton in unterschiedlichem Umfang sowie mit unterschiedlicher Verbindlichkeit und Zuständigkeit. Der Kanton Wallis führt in beiden Sprachregionen kantonale Prüfungen durch. Die Praxis in der Westschweiz variiert stark zwischen den Kantonen. Es werden zertifizierende und formative Prüfungen sowie Prüfungen im Rahmen des kantonalen Monitorings durchgeführt. Die Kantone Genf, Jura, Neuenburg, Waadt und Wallis setzen jeweils mehrere Prüfungen in verschiedenen Schulfächern ein. Seit 2016 entwickelt das *Institut de recherche et documentation pédagogique* (IRDP) im Auftrag der *Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin* (CIIP) im Rahmen des Projekts EpRoCom eine Westschweizer Itembank mit Testaufgaben, die sich auf die Lernziele und Progressionen des Westschweizer Lehrplans (*Plan d'études romand* PER) beziehen und den Kantonen und den Lehrpersonen zur Verfügung gestellt werden sollen. In einer ersten Etappe wurden aus Prüfungen der Westschweizer Kantone Aufgaben für Mathematik und Französisch im 8. Schuljahr ausgewählt und einem Validierungsprozess unterworfen. Seit 2020 existiert ein bereinigtes Gesamtkonzept für das Projekt *EpRoCom – Banque d'items*. Im Jahr 2021 ist eine Auswahl von Aufgaben auf der Plattform des PER den Lehrpersonen vor allem zu Illustrationszwecken zur Verfügung gestellt worden. Der Kanton Tessin verfügt in in einer Reihe von Fächern über kantonale Prüfungen und baut ein Test- und Monitoringsystem auf, das nationale und kantonale Erhebungen integriert. Individuelle Standortbestimmungen mit dem Ziel der Förderplanung wurden bisher nicht eingeführt. Am Ende der Primarstufe und der Sekundarstufe I werden standardisierte kantonale Prüfungen durchgeführt (Schwerpunkt Mathematik und Italienisch), die als Rückmeldung an die Lehrpersonen dienen. Weitere kantonale Prüfungen mit Bezug auf den Tessiner Lehrplan werden nur auf der Sekundarstufe I durchgeführt. Bisher erfolgen die Auswahl der Schuljahre und Fächer, die getestet werden, sowie die Stichprobenziehung noch nicht nach einem festen Konzept.

Das Informations- und Dokumentationszentrum IDES der EDK erfasst in seinen Kantonsumfragen alljährlich, welche externen, für alle Schülerinnen und Schüler verbindlichen Prüfungen in den Schweizer Kantonen und im Fürstentum Liechtenstein eingesetzt werden (EDK/IDES, 2021).

1.1.5 Koordinationsstab HarmoS

Der Umsetzungsbeschluss zum HarmoS-Konkordat markiert im Bereich der Harmonisierung den Übergang von einer Projektstruktur (HarmoS-Projekt) in die permanenten Strukturen der EDK. Im Verlauf des Jahres 2008 nimmt der Koordinationsstab für die Umsetzung der Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS) seine Tätigkeit auf. Dabei handelt es sich um eine ständige Kommission der EDK gemäss Art. 21 des EDK-Statuts. Aufgabe des Kosta HarmoS ist es, den koordinierten Vollzug des HarmoS-Konkordats in der Sache und zwischen den interkantonalen Umsetzungsebenen zu gewährleisten. Seine Zusammensetzung, Aufgaben und Arbeitsweise sind im

[5] Das Fürstentum Liechtenstein führt obligatorische Überprüfungen seiner Bildungsstandards in den Fächern Deutsch und Mathematik durch (5., 7. und 10. Stufe, zusätzlich Englisch auf Stufe 7 und 10) (EDK/IDES, 2021).

Reglement für den Kosta HarmoS vom 24. Januar 2008 (EDK, 2013b) sowie im Organisationsreglement für die Durchführung der ÜGK festgelegt (ÜGK-Reglement; EDK, 2014b).

1.2 Umsetzung der ÜGK

1.2.1 Rechtsgrundlagen zur Umsetzung der ÜGK

Am 20. Juni 2013 verabschiedet die EDK-Plenarversammlung das Konzept für die Durchführung der ersten beiden Erhebungen im Rahmen der ÜGK und beschliesst die ersten ÜGK-Erhebungen (EDK, 2013a). Am 8. Mai 2014 folgt der Erlass des ÜGK-Reglements.

Gemäss dem Beschluss vom 20. Juni 2013 soll im Jahr 2016 im 11. Schuljahr (gem. HarmoS) das Erreichen der Grundkompetenzen in Mathematik getestet werden, im Jahr 2017 im 8. Schuljahr das Erreichen der Grundkompetenzen in der Schulsprache und in der ersten Fremdsprache. Zudem wird in den Erwägungen für 2019 ein ÜGK-Zyklus mit den Fachbereichen Mathematik und Naturwissenschaften im 4. Schuljahr in Aussicht gestellt, der allerdings nicht realisiert wird.

Das Konzept zur ÜGK umreisst insbesondere die Organisationsstruktur, einen Zeitplan für die Erhebungen, Art und Umfang der Stichproben von Schülerinnen und Schülern, und legt zudem eine Finanzplanung vor. Das GS EDK übernimmt demnach die Funktion einer Koordinationsstelle, während die ADB und ein noch zu bildendes wissenschaftliches Konsortium arbeitsteilig für die Ausführung der Arbeiten zuständig sind (Weiteres unter 1.2.3). Der Zeitplan für künftige ÜGK-Erhebungen zeigt auf, dass ab der ÜGK 2017 die Möglichkeit bestehen würde, Längsschnitte auf der Ebene der Individuen zu bilden, um Entwicklungen der Schülerinnen und Schüler innerhalb der Schulzeit verfolgen zu können. Bei der Umsetzung der ÜGK verzichtet man darauf, konsequent Längsschnittdaten auf der Ebene des Individuums zu erheben, und beschränkt man sich auf Kohortenlängsschnitte, also Erhebungen bei Stichproben aus demselben Jahrgang. Demnach wären die Sprachkompetenzen 2017 im 8. und 2020 im 11. Schuljahr getestet worden, wenn die Corona-Pandemie dies nicht im letzten Moment verhindert hätte.

Gemäss ÜGK-Reglement besteht das Ziel der ÜGK darin, die Leistungen des Bildungssystems der obligatorischen Schule zu evaluieren und den Kantonen auf der Systemebene Informationen darüber zu liefern, inwieweit die Grundkompetenzen auf bestimmten Schulstufen erreicht werden. Im Vordergrund stehen demnach die Testergebnisse gemessen an den Bildungszielen. Die Erhebung von Kontextdaten (z.B. zum sozialen Hintergrund der Schülerinnen und Schüler), die der Interpretation der Ergebnisse dienen können, wird im ÜGK-Reglement nicht erwähnt.

Das ÜGK-Reglement regelt u.a. auch die Organisation, die Zuständigkeiten und die Steuerung der ÜGK durch die EDK, insbesondere

- die Mitwirkung der Kantone bei den Erhebungen;
- die Zuständigkeiten der Plenarversammlung (u.a. Freigabe der Ergebnisse der ÜGK für die Berichterstattung, Genehmigung der Festlegung des Schwellenwerts in Übereinstimmung mit den Grundkompetenzen);
- die Aufgaben des Kosta HarmoS im Zusammenhang mit der ÜGK;
- Qualifikationen und Zuständigkeiten des wissenschaftlichen Konsortiums, das mit der Durchführung der ÜGK betraut wird;
- die Datensicherheit bzw. den Umgang mit den erhobenen Daten.

Nach dem Beschluss zu den ersten Erhebungen im Jahr 2013 stimmt die Plenarversammlung der EDK am 22. Juni 2017 (EDK, 2017) der Durchführung von zwei weiteren Erhebungen im Rahmen der ÜGK zu. Demnach soll im Jahr 2020 im 11. Schuljahr das Erreichen der Grundkompetenzen in den Fachbereichen Schulsprache sowie erste und zweite Fremdsprache überprüft werden, im Jahr 2022 im 4. Schuljahr in zwei weiteren Fachbereichen. Am 25. Oktober 2019 (EDK, 2019c) beschliesst die EDK-Plenarversammlung, dass es sich dabei um Schulsprache und Mathematik handeln soll. Aufgrund der

Pandemie muss die ÜGK 2020 aber gestoppt werden, und die EDK-Plenarversammlung beschliesst am 25. Juni 2020 einen neuen Zeitplan: Die ÜGK im 11. Schuljahr findet demnach im Jahr 2023, diejenige im 4. Schuljahr 2024 statt (EDK, 2020b).

1.2.2 Aufgabendatenbank

Am 25. Oktober 2012 beschliesst die EDK-Plenarversammlung die Schaffung der Aufgabendatenbank (ADB) der EDK (EDK, 2012b), und am 23. Januar 2014 erlässt der Vorstand der EDK das Organisationsreglement für die Aufgabendatenbank (ADB-Reglement) (EDK, 2014a). Die ADB erhält gemäss Reglement eine Reihe von Aufgaben, die für die Umsetzung der Art. 8 und 10 des HarmoS-Konkordats (regelmässige Referenztests) von zentraler Bedeutung sind:

a) eine multifunktionale und mehrsprachige technische Struktur bereitzustellen;
b) eine gesamtschweizerische Messskala sicherzustellen;
c) einen Aufgaben- und Testentwicklungsprozess zu etablieren, in dessen Rahmen Aufgaben sowie Tests entwickelt und an den Grundkompetenzen geeicht werden können;
d) den Austausch von Aufgaben für die Überprüfung der Erreichung der nationalen Bildungsziele (Grundkompetenzen) sowie den Austausch von Aufgaben für Projekte der Kantone zu koordinieren.

Die ADB ist als «Hüterin der Grundkompetenzen» (Eichung von Aufgaben an den Grundkompetenzen; Sicherstellung einer gesamtschweizerischen Messskala) anders positioniert als etwa die Institutionen, die sich später aufgrund von Vereinbarungen mit der EDK an der Durchführung der ÜGK beteiligen. Die Verantwortung für die ADB liegt direkt bei der EDK. Diese überträgt die strategischen und operativen Aufgaben an den Kosta HarmoS und an die Geschäftsstelle ADB (GS ADB). Die GS ADB ist organisatorisch bei der Schweizerischen Koordinationsstelle für Bildungsforschung (SKBF) angesiedelt. Mit Beschluss vom 26. Oktober 2018 ist der Betrieb der ADB bis vorläufig 2025 bewilligt.

In der Umsetzung nimmt die ADB letztlich ein weniger breites Spektrum an Aufgaben wahr, als im Reglement vorgesehen. Mit dem Aufbau einer multifunktionalen Aufgabendatenbank im Sinne von a) wird zwar begonnen, ihr Unterhalt wird aber eingestellt, vor allem deshalb, weil deren Möglichkeiten von Regionen und Kantonen nicht nachgefragt werden. Als technische Plattform verwendet man in einer ersten Phase das E-Testing-System TAO, das auch bei der ÜGK 2016 (11. Schuljahr) eingesetzt wird. Bei dieser Erhebung wird TAO internetbasiert auf den schuleigenen Infrastrukturen eingesetzt; alternativ dazu könnte TAO auch von USB-Sticks aus eingesetzt werden. Bei der ÜGK 2017 (8. Schuljahr) geht man dazu über, für die Datenerhebung eigene Geräte (Tablets in Koffersets) in die Schulen zu bringen. Der Hauptgrund dafür sind die weniger ausgebauten Computerinfrastrukturen an den Primarschulen. Mit der Bring-in-Lösung wird auch die Testsoftware gewechselt. Neu wird TBA ItemBuilder des DIPF (Leibniz-Institut für Bildungsinformation und Bildungsforschung) eingesetzt. Die Mitarbeitenden der HTW Chur (heute Teil der Fachhochschule Graubünden FHGR), die mit der technischen Umsetzung der ÜGK betraut sind, sowie die Aufgabenentwicklerinnen und Aufgabenentwickler der Fremdsprachentests sind mit der Software aus anderen Kontexten bereits vertraut. Ein Austausch von Aufgaben im Sinne von d) kommt nur ganz rudimentär in Gang, und zwar in Form einer Freigabe eines Teils der Aufgaben aus den ÜGK-Zyklen 2016 und 2017. Diese können, ähnlich wie die Daten der ÜGK, beim Kompetenzzentrum für Sozialwissenschaften FORS an der Universität Lausanne bezogen werden, was *Stellwerk* auch getan hat.

Die bisherigen Arbeiten der ADB konzentrieren sich auf die Aufgabenbereiche b) und c). Dabei handelt es sich um Arbeiten im Zusammenhang mit den vergangenen ÜGK-Zyklen 2016 und 2017 sowie die abgeschlossenen Vorbereitungsarbeiten im Hinblick auf die ÜGK 2023 (ursprünglich 2020) und die noch laufende Vorbereitung der ÜGK 2024 (ursprünglich 2022). Für alle vier Erhebungen hat die ADB zusammen mit Gruppen von Fachdidaktikerinnen und Fachdidaktikern Aufgaben und Tests zur Überprüfung des Erreichens der Grundkompetenzen entwickelt, Testdaten von Pilot- und Haupterhebungen analysiert sowie bei den vergangenen ÜGK-Zyklen die Schwellenwertsetzung geleitet und zur Berichterstattung beigetragen.

1.2.3 Wissenschaftliches Konsortium, Projektleitung ÜGK und ÜGK-Ausschuss

Die EDK arbeitet schon früh darauf hin, dass ein wissenschaftliches Konsortium unter der Leitung einer universitären Hochschule die wissenschaftliche Projektleitung für die ÜGK übernimmt und somit die Durchführung der Erhebungen koordiniert und die Einbindung der ÜGK in die Wissenschaft sicherstellt. Dies ist in Art 7 des ÜGK-Reglements vom 8. Mai 2014 verankert.

Um ein wissenschaftliches Konsortium aufzubauen, werden für Interessierte ab 2012 mehrere runde Tische durchgeführt. Das Ziel dieser runden Tische besteht darin, gemeinsam mit Expertinnen und Experten der Leistungsmessung die Kriterien für eine nachhaltige und erfolgreiche Institutionalisierung der Projektstrukturen an einer oder mehreren Hochschulen unter Berücksichtigung der sprachregionalen Bedürfnisse zu erarbeiten. In dieser Phase erweist sich der Mangel an Erfahrung mit eigenen Large-Scale-Assessments in der Schweiz als Schwierigkeit. Weiter führt die knappe Finanzierung dazu, dass es lange nicht gelingt, eine funktionierende wissenschaftliche Struktur als Trägerin der ÜGK zu schaffen.

Im Jahr 2014 gelingt es nicht rechtzeitig für die ersten beiden ÜGK-Zyklen, ein wissenschaftliches Konsortium zu bilden, das die wissenschaftliche Verantwortung für die Durchführung übernehmen würde. Aus diesem Grund ruft das Generalsekretariat der EDK eine Projektleitung ÜGK (PL ÜGK) ins Leben, welche die operativen Geschäfte bis zur Einsetzung des wissenschaftlichen Konsortiums führt. Neben dem GS EDK sind darin die GS ADB, die regionalen Durchführungszentren, die bereits die PISA-Studie durchführen, (Service de la recherche en éducation SRED, Pädagogische Hochschule St. Gallen PHSG, Scuola universitaria professionale della Svizzera Italiana SUPSI), als Verantwortliche für die Technik die Hochschule für Technik und Wirtschaft HTW Chur (heute Teil der FHGR), das für die Stichprobenziehung verantwortliche Institut für Bildungsevaluation IBE (assoziiertes Institut der Universität Zürich) und das für die Datenaufbereitung und -bereitstellung beauftragte Kompetenzzentrum für Sozialwissenschaften FORS an der Universität Lausanne vertreten. Ebenfalls Teil der PL ÜGK ist die Leitung des Aufbaugremiums für das zukünftige wissenschaftliche Konsortium.

Um diese operativen Strukturen auf der strategischen Ebene angemessen zu ergänzen, setzt der Kosta HarmoS an seiner Sitzung vom 21. November 2014 aus den eigenen Reihen den ÜGK-Ausschuss ein, der als repräsentatives Element der Koordination und Steuerung gegenüber den wissenschaftlichen Institutionen handelt. In ihm sind alle Sprachregionen sowie der Bildungsraum Nordwestschweiz vertreten.

Das Aufbaugremium, bestehend aus Vertreterinnen und Vertretern der Universität Bern (Leitung), der Universität Genf und der Scuola universitaria professionale della Svizzera italiana (SUPSI), arbeitet bis Ende 2016 im Einvernehmen mit anderen Institutionen eine Geschäftsordnung des wissenschaftlichen Konsortiums Bildungsforschung und -evaluation (WIKO BFE) aus. Der Kosta HarmoS nimmt diese am 27. Januar 2017 zustimmend zur Kenntnis. Der Entwurf der Geschäftsordnung vom 7. November 2016 sieht vor, dass sich das WIKO BFE mit der Koordination und Durchführung von Projekten von nationalem Interesse im Bereich Bildungsforschung und -evaluation befasst, besondere auch mit PISA und der ÜGK, die nachhaltige wissenschaftliche Einbettung solcher Projekte fördert, diese im Netzwerk der Wissenschaft verankert und als Ansprech- und Diskussionspartner der Bildungsbehörden fungiert. Als Organe sind

a) ein wissenschaftlicher Beirat mit Beratungs-, Leitungs- und Entscheidungsfunktionen (Universitäten Bern und Genf, SUPSI; evtl. weitere Mitglieder in beratender Funktion),
b) eine Professur für Bildungsforschung und -evaluation an der Universität Bern,
c) eine Projektleitung,
d) eine Delegiertenversammlung mit beratender Funktion von fachlich interessierten Vertreterinnen und Vertretern von Schweizer Hochschulen sowie
e) ein internationaler Rat von drei Expertinnen und Experten aus verschiedenen Sprachräumen vorgesehen.

Im Hinblick auf die Übernahme der wissenschaftlichen Projektleitung für die ÜGK-Erhebungen 2020 und 2022 bzw. die Koordination des wissenschaftlichen Konsortiums schliessen die EDK und die Universität Bern im September 2018 eine Kooperationsvereinbarung ab. Das GS EDK zieht sich damit aus der

Projektleitung zurück. Die Leitung der ÜGK wird (wie später auch diejenige für PISA 2021 bzw. 2022) am Interfaculty Centre for Educational Research (ICER) angesiedelt. Das ICER der Universität Bern arbeitet im Rahmen der PL ÜGK mit denselben Partnern zusammen wie zuvor das GS EDK, ausser mit FORS. FORS ist zwar nicht mehr Teil des Projekts, stellt aber im Rahmen seines Dienstleistungsauftrags die Daten der ÜGK auf seinen Plattformen FORSbase bzw. SWISSUbase (seit 2022) zur Verfügung. In der PL ÜGK sind ICER, ADB, FHGR, IBE, PHSG, SRED und SUPSI vertreten. Das ICER (bzw. die Universität Bern) schliesst mit allen Partnern ausser der ADB und teilweise der FHGR eigene Vereinbarungen ab. Die ADB ist Teil der EDK; mit der FHGR hat die EDK im Sinne einer Übergangslösung eine separate Kooperationsvereinbarung im Hinblick auf Informatikdienstleistungen im Zusammenhang mit der ÜGK 2020 bzw. 2023. Die Kooperationsvereinbarung zwischen EDK und Universität Bern sieht für die Universität Bern (und damit das ICER) in der ÜGK zum einen Koordinations- und zum anderen Leitungsaufgaben vor. Wie im Prozesshandbuch des GS EDK zur Durchführung von Large-Scale-Erhebungen vorgesehen, verantwortet eine Reihe von thematischen Arbeitsgruppen die anfallenden Arbeiten: AG Testentwicklung, AG Stichprobenziehung, AG Kontextfragebogen, AG Testauslieferung, AG Erhebungsmanagement, AG Datenaufbereitung, AG Archivierung und Bereitstellung, AG Schwellenwertsetzung, AG Erstberichterstattung.

Die neue Struktur gemäss Kooperationsvereinbarung zwischen der EDK und der Universität Bern ist keine direkte Umsetzung des Entwurfs der Geschäftsordnung für das WIKO BFE. Insbesondere liegt die wissenschaftliche Verantwortung viel stärker als dort skizziert bei nur einer Hochschule, der Universität Bern. Diese stellt auch den Vorstand des ICER, der Funktionen des interinstitutionell angedachten wissenschaftlichen Beirats und der im Entwurf zur Geschäftsordnung unter b) genannten Professur wahrnimmt. Eine spezialisierte Professur ist an der Universität Bern bisher nicht eingerichtet worden. Aufgaben der Projektleitung werden hauptsächlich von der Direktion des ICER wahrgenommen, die Teil der Projektleitung der ÜGK ist und diese koordiniert und teilweise leitet. Die unter d) aufgeführte Delegiertenversammlung von interessierten Hochschulen existiert nicht. Das ICER hat sich aber verpflichtet, diese Zielgruppe periodisch zu runden Tischen einzuladen. Zudem haben interessierte Forschende unter bestimmten Bedingungen die Möglichkeit, ihre Forschungsinteressen in den Kontextfragebogen zur ÜGK einzubringen. Der internationale Rat von Expertinnen und Experten ist nach anderen Prinzipien zusammengesetzt als unter e) beschrieben und ist ein Gremium des ICER, d.h. es befasst auch mit Vorhaben des ICER, die nicht spezifisch nationale Interessen im Bereich der Bildungsforschung und -evaluation betreffen.

1.2.4 Zusammenarbeit mit der Fachdidaktik

Wie in Abschnitt 1.1.2 beschrieben, sind Vertreterinnen und Vertreter der Fachdidaktiken der vier Fachbereiche, zu denen Bildungsstandards entwickelt wurden, bereits früh aktiv in die Arbeiten der EDK involviert. In den Strukturen der ÜGK sind die Fachdidaktiken jedoch nicht ständig vertreten. Vielmehr schliesst die EDK immer dann, wenn das Erreichen der Grundkompetenzen in einem Fachbereich überprüft werden soll, mit den Institutionen interessierter Fachdidaktikerinnen und Fachdidaktiker eine Kooperationsvereinbarung für die Aufgaben- und Testentwicklung sowie die Beteiligung an der Schwellenwertsetzung ab. In den Fachdidaktik-Konsortien sind in der Regel die drei grösseren Sprachregionen vertreten. Eine Ausnahme stellen die Fremdsprachen in der ÜGK 2017 dar. Da leistet das Institut für Mehrsprachigkeit (IFM) der Universität Freiburg und der Pädagogischen Hochschule Freiburg allein die Aufgabenentwicklung. Die Fachdidaktikkonsortien organisieren ihre Zusammenarbeit in Absprache mit der GS ADB und dem GS EDK selbst. Fachdidaktikerinnen und Fachdidaktiker haben zudem die Möglichkeit, zum Kontextfragebogen beizutragen. Diesen Beitrag leisten sie ausgehend von den runden Tischen des WIKO ausserhalb der Vereinbarung mit der EDK. Einzelne Vertreterinnen und Vertreter der Fachdidaktik tragen auch zur Erstberichterstattung bei. Dafür werden separate Vereinbarungen abgeschlossen.

1.2.5 Das Netzwerk der kantonalen Referenzpersonen

Um die Erhebungen in den Kantonen und insbesondere auch den Kontakt mit den teilnehmenden Schulen zu koordinieren, beschliesst der Kosta HarmoS am 10. April 2014, ein Netzwerk von kantonalen Referenzpersonen der ÜGK zu bilden. Die Kantone sind darin in der Regel durch eine Person vertreten, mehrsprachige Kantone durch eine Person pro Sprachregion. Die Referenzpersonen initiieren im Auftrag ihrer kantonalen Departemente gegenüber den Schulen die Erhebungen, indem sie den Kontakt zwischen den Schulen und den regionalen Durchführungszentren herstellen. Sie bleiben während des ganzen Prozesses sowohl für die Schulen als auch für die Durchführungszentren für Fragen der Durchführung der Erhebungen ansprechbar. Seit der Kick-off-Veranstaltung am 18. November 2014 finden jährlich in der Regel zwei bis drei Treffen der kantonalen Referenzpersonen statt, bei denen über den Stand der Arbeiten informiert wird und ein Austausch über verschiedene Planungs-, Kommunikations- und Durchführungsfragen stattfindet. Am 24. Januar 2019 beschliesst der EDK-Vorstand ein Mandat für die Arbeitsgruppe der kantonalen Referenzpersonen für Large-Scale-Assessments in der Schweiz.

1.3 Durchgeführte ÜGK-Erhebungen

Im Folgenden werden die wichtigsten Eckwerte der Erhebungen 2016 und 2017 beschrieben. Die beiden ersten ÜGK-Zyklen finden zu einem Zeitpunkt statt, zu dem sich die neuen, auf die Bildungsziele bezogenen Lehrpläne erst in der Umsetzung befinden. Es handelt sich also um Initialmessungen, die eine erste Baseline liefern sollen.

Die Darstellung folgt weitgehend dem ausführlichen «Bericht zu den Erhebungen 2016 und 2017» (EDK, 2019a). Im Zusammenhang mit der ÜGK 2016 wird auf den Auditbericht (Fischbach & Ugen, 2018) eingegangen, der dazu in Auftrag gegeben wurde, und es wird ein Überblick über die Berichterstattung, die Bereitstellung der Daten und den Aufwand für die beiden ersten ÜGK-Zyklen gegeben.

1.3.1 ÜGK 2016: Mathematik im 11. Schuljahr

Die erste ÜGK-Erhebung findet im Frühjahr 2016 statt. Dabei wird bei Schülerinnen und Schülern am Ende der obligatorischen Schulzeit (11. Schuljahr) das Erreichen der Grundkompetenzen in Mathematik überprüft. Insgesamt nehmen 22‘423 Schülerinnen und Schüler aus allen Kantonen der Schweiz teil. Für die Durchführung der ÜGK 2016 verantwortlich ist die in Abschnitt 1.2.3 beschriebene interimistische Projektleitung.

Aufgaben- und Testentwicklung

Die Grundlage für die Aufgaben- und Testentwicklung der ÜGK 2016 bilden die nationalen Bildungsziele für die Mathematik, die 2011 von der EDK freigegeben wurden. Für die ÜGK 2016 werden Testaufgaben für alle fünf Kompetenzbereiche sowie fünf von acht Handlungsaspekten des Kompetenzmodells entwickelt. Fachdidaktikerinnen und Fachdidaktiker von pädagogischen Hochschulen und Universitäten aus drei Sprachregionen der Schweiz (d, f, i) erarbeiten die Aufgaben in Zusammenarbeit mit der GS ADB. Die Aufgaben durchlaufen mehrere Review- und Validierungsprozesse und werden im Frühjahr 2015 an einer national repräsentativen Stichprobe pilotiert. Anlässlich der Pilotierung zeigt sich, dass viele Items für die Schülerinnen und Schüler eher schwierig sind. Zudem weisen einige Items für die Schülerinnen und Schüler in verschiedenen Sprachregionen unterschiedliche Lösungswahrscheinlichkeiten auf (*differential item functioning* DIF), was hauptsächlich auf inkonsistente Übersetzungen zurückzuführen ist. Im Anschluss an die Pilotierung werden deshalb die Übersetzungen überarbeitet, und es werden zusätzliche Testaufgaben für den unteren Fähigkeitsbereich entwickelt. Bei der Haupterhebung werden somit auch Aufgaben eingesetzt, die nicht oder nicht in ihrer finalen Version in der Pilotierung erprobt wurden (vgl. dazu Angelone & Keller, 2019a).

Kontextfragebogen

Um in erster Linie Gründe und Ursachen für Leistungen bzw. Leistungsunterschiede identifizieren zu können, werden mit einem Fragebogen für die Schülerinnen und Schüler verschiedene Kontextinformationen erhoben. Der Kontextfragebogen enthält insbesondere Fragen zur sozialen Herkunft, zu den zuhause gesprochenen Sprachen, zum Migrationshintergrund, zu Emotionen gegenüber der Schule, zu Einstellungen zur Mathematik und zur Motivation für das Fach sowie zu Aspekten der wahrgenommenen Qualität des Mathematikunterrichts. Für die ÜGK 2016 werden zwei Fragebogenversionen entwickelt, die eine Überlappung im Bereich des sozialen Hintergrunds aufweisen, sonst aber unterschiedliche Aspekte erfassen. Die beiden Versionen werden zufällig auf die individuellen Schülerinnen und Schüler verteilt.

Durchführung

Die Haupterhebung der ÜGK 2016 findet zwischen dem 2. Mai und dem 10. Juni 2016 statt und erfolgt nach einem standardisierten Vorgehen. Geschulte Testverantwortliche sind vor Ort und leiten die Erhebungen. Die Schülerinnen und Schüler bearbeiten die Kompetenztest und Fragebogen online an den Computern ihrer Schule. Die Erhebung dauert insgesamt 3 Stunden und 15 Minuten.

Schwellenwertsetzung

Der Schwellenwert, ab dem die Grundkompetenzen als erreicht gelten, wird in einem Standardsetting-Verfahren von einer Expertengruppe, bestehend aus Fachdidaktikerinnen und Fachdidaktikern sowie Lehrpersonen aus drei Sprachregionen der Schweiz (d, f, i) anhand einer modifizierten Bookmark-Methode festgelegt. Ein Teil der Fachdidaktikerinnen und Fachdidaktiker der Expertengruppe hat an der Aufgaben-entwicklung mitgewirkt, andere nicht. Die Auswahl der Mitglieder der Expertengruppe ist durch die Geschäftsstellen der EDK-Regionalkonferenzen erfolgt (vgl. Angelone & Keller, 2019a).

Ergebnisse

Das zentrale Ergebnis der ÜGK 2016 ist, dass rund zwei Drittel der Schülerinnen und Schüler am Ende der obligatorischen Schule die Grundkompetenzen in Mathematik erreichen. Allerdings bestehen zwischen den Kantonen vergleichsweise grosse Unterschiede, die sich nicht allein durch die Zusammensetzung der Schülerschaft erklären lassen. So erreichen beispielsweise in den französischsprachigen Gebieten der Kantone Freiburg und Wallis mehr als 80 Prozent der Schülerinnen und Schüler die Grundkompetenzen in Mathematik, während im Kanton Basel-Stadt nur 44 Prozent und in den Kantonen Bern (deutschsprachiger Teil), Basel-Landschaft, Luzern und Solothurn jeweils nur rund 55 Prozent der Schülerinnen und Schüler die Grundkompetenzen erreichen (vgl. EDK, 2019b; Nidegger, Roos, Petrucci, Verner, Helbling, Arnold, Erzinger, Pham, Schönenberger, et al., 2019).

1.3.2 ÜGK 2017: Schulsprache und erste Fremdsprache im 8. Schuljahr

Die zweite ÜGK-Erhebung findet im Frühjahr 2017 statt. Bei Schülerinnen und Schülern im 8. Schuljahr wird das Erreichen der Grundkompetenzen in der Schulsprache (Deutsch, Französisch oder Italienisch) und in der ersten Fremdsprache (zweite Landessprache oder Englisch) überprüft. An der Haupterhebung der ÜGK 2017 nehmen insgesamt 20'177 Schülerinnen und Schüler aus allen Kantonen der Schweiz teil. Der Kanton Graubünden beteiligt sich wegen seiner besonderen Situation nicht an den Tests im Fachbereich Fremdsprachen.

Aufgaben- und Testentwicklung

Die Testaufgaben für die Schulsprache werden von Fachdidaktikerinnen und Fachdidaktikern von pädagogischen Hochschulen aus drei Sprachregionen (PH FHNW, HEP Vaud, SUPSI) in Zusammenarbeit mit der GS ADB entwickelt. Getestet wird das Erreichen der Grundkompetenzen in den Teilbereichen Lesen und Orthografie. Die Aufgaben durchlaufen mehrere Review- und Validierungszyklen und werden im Frühjahr 2016 an einer national repräsentativen Stichprobe pilotiert (vgl. Angelone & Keller, 2019b).

Für die erste Fremdsprache entwickelt das Institut für Mehrsprachigkeit (IFM) der Universität Freiburg und der Pädagogischen Hochschule Freiburg in Zusammenarbeit mit der GS ADB die Testaufgaben. Es werden die Kompetenzbereiche Leseverstehen und Hörverstehen getestet. Die erste Fremdsprache ist in der Romandie Deutsch, im Tessin und in den Deutschschweizer Kantonen, die an die Romandie angrenzen, Französisch (Region des ehemaligen «Passepartout»-Lehrplans), in den übrigen Kantonen der Deutschschweiz Englisch. Die Aufgaben durchlaufen mehrere Reviewprozesse und werden zuerst in drei Sprachregionen (d, f, i) vorpilotiert und dann im Frühjahr 2016 an einer national repräsentativen Stichprobe pilotiert (vgl. Angelone & Keller, 2019b).

Kontextfragebogen

Für die ÜGK 2017 entwickeln Expertinnen und Experten aus Institutionen, die in der PL ÜGK oder in der Aufgabenentwicklung vertreten sind, zusammen mit Forschenden mit einem spezifischen Interesse, die sonst nicht an der ÜGK beteiligt sind, den Kontextfragebogen. Er wird im Jahr 2016 im Rahmen der nationalen Piloterhebung miterprobt. Mit dem Fragebogen wird neben der sozialen Herkunft, dem Migrationshintergrund und den zuhause gesprochenen Sprachen eine Reihe von weiteren Aspekten erhoben, vorwiegend solche, die bezüglich der schulisch erworbenen Sprachkompetenzen als besonders relevant erachtet werden, beispielsweise solche zum Sprachenlernen, zu Sprachkontakten, zur Sprachlernmotivation, zu Emotionen im Zusammenhang mit Sprachen und zur Wahrnehmung der Unterrichtsgestaltung (vgl. Erzinger et al., 2019).

Durchführung

Die Haupterhebung der ÜGK 2017 findet zwischen dem 24. April und dem 2. Juni 2017 statt. Wiederum führen geschulte Testverantwortliche die Testsitzungen gemäss einem standardisierten Verfahren durch. Die Schülerinnen und Schüler bearbeiten, anders als in der ÜGK 2016, die Kompetenztests und Fragebogen nun auf Tablets, die von den Testverantwortlichen in die Schulen mitgebracht worden sind. Die Erhebung dauert 2 Stunden und 55 Minuten.

Schwellenwertsetzung

Das Verfahren der Schwellenwertsetzung bei den Schul- und Fremdsprachen entspricht dem Verfahren bei der Schwellenwertsetzung für die Mathematik. Wiederum wird unter der Leitung der ADB nach einer modifizierten Bookmark-Methode vorgegangen. Die Expertengruppen setzen sich aus Fachdidaktikerinnen und Fachdidaktikern, die an der Aufgabenentwicklung beteiligt waren, nicht beteiligten Fachdidaktikerinnen und Fachdidaktikern, praktizierenden Lehrpersonen der Primarstufe sowie Heilpädagoginnen zusammen (vgl. Angelone & Keller, 2019b).

Ergebnisse

Die zentralen Ergebnisse der ÜGK 2017 lassen sich wie folgt zusammenfassen (vgl. EDK, 2019b; Nidegger, Roos, Petrucci, Verner, Helbling, Arnold, Erzinger, Pham, Ambrosetti, et al., 2019):

- Im Lesen in der Schulsprache erreichen rund 88% der Schülerinnen und Schüler die Grundkompetenzen. Dabei erreichen im französischsprachigen Teil des Kantons Freiburg mit 91.3% statistisch signifikant mehr Schülerinnen und Schüler die Grundkompetenzen als im schweizerischen Durchschnitt. Statistisch signifikant unter dem Schweizer Mittelwert liegen die Anteile in den Kantonen Basel-Stadt (80.9%) und Glarus (80.5%).
- In der Orthografie in der Schulsprache Deutsch erreichen durchschnittlich rund 84% der Schülerinnen und Schüler die Grundkompetenzen, in der Orthografie in der Schulsprache Französisch rund 89% und in der Orthografie in der Schulsprache Italienisch rund 80%.
- In der Fremdsprache Deutsch (Westschweiz) erreichen im Hörverstehen im Durchschnitt 88% der Schülerinnen und Schüler die Grundkompetenzen, im Leseverstehen rund 72%. Die Unterschiede zwischen den Kantonen sind gering. Einzig im französischsprachigen Teil des Kantons Freiburg

erreicht im Leseverstehen ein statistisch signifikant höherer Anteil der Lernenden die Grundkompetenzen.
- In der Fremdsprache Französisch («Passepartout»-Region der Deutschschweiz und Tessin) erreichen im Hörverstehen im Durchschnitt rund 89% der Schülerinnen und Schüler die Grundkompetenzen, im Leseverstehen rund 65%. Zwischen den Kantonen bestehen allerdings beträchtliche Unterschiede. Im Kanton Tessin sowie im deutschsprachigen Teil des Kantons Freiburg erreichen jeweils statistisch signifikant mehr Schülerinnen und Schüler die Grundkompetenzen, in den Kantonen Basel-Stadt, Basel-Landschaft und Solothurn statistisch signifikant weniger.
- In der Fremdsprache Englisch erreichen im Hörverstehen im Durchschnitt rund 95% der Schülerinnen und Schüler die Grundkompetenzen, im Leseverstehen rund 86%. Die Unterschiede zwischen den Kantonen sind gering. Nur in den Kantonen Schaffhausen (Leseverstehen und Hörverstehen) und im Kanton Appenzell Innerrhoden (Leseverstehen) erreichen statistisch signifikant mehr Schülerinnen und Schüler die Grundkompetenzen.

Die Analysen zeigen zudem, dass die unterschiedlichen Anteile beim Erreichen der Grundkompetenzen zwischen den Kantonen nicht – oder nur zu einem äusserst geringen Teil – auf die unterschiedlichen kantonalen Schülerzusammensetzungen zurückgeführt werden können.

1.3.3 Auditbericht zur ÜGK 2016

Am 27. Oktober 2017 wird die Plenarversammlung über erste ÜGK-Ergebnisse im Bereich Mathematik informiert. Das GS EDK erhält sodann den Auftrag, weitere Abklärungen zu treffen, die insbesondere die wissenschaftliche Eignung der Verfahren zur Aufgabenentwicklung und Schwellenwertsetzung in den Blick nehmen sollen. Im November 2017 werden Dr. Antoine Fischbach und Dr. Sonja Ugen, Experte bzw. Expertin des Luxembourg Centre for Educational Testing (LUCET) der Universität Luxemburg, mit dem Verfassen eines Auditberichts beauftragt. Die zentralen Fragen, die mit dem Bericht beantwortet werden sollen, sind die folgenden:

- Entsprechen die Prozesse der Aufgaben- und Testentwicklung und der Schwellenwertsetzung den üblichen wissenschaftlichen Anforderungen?
- Werden mit den ÜGK-Aufgaben die von der EDK freigegebenen Grundkompetenzen gemessen?
  a) Bezüglich der Kompetenzbereiche und der Handlungsaspekte?
  b) Bezüglich der Schwierigkeit?

Der Autor und die Autorin des Auditberichts kommen zu folgenden Ergebnissen (Fischbach & Ugen, 2018):

- Die Aufgabenentwicklung sollte verbessert werden. Die Aufgabenentwicklungsprozesse entsprechen nicht dem State-of-the-Art und sind das schwächste Glied in der ÜGK 2016. Dennoch ist der letztlich eingesetzte Test akzeptabel, und die daraus hervorgegangenen Daten sind plausibel und interpretierbar.
- Die Schwellenwertsetzung sowie das Testdesign, die Datenerhebung und die Datenanalyse liegen auf einem sehr hohen Niveau und entsprechen den internationalen wissenschaftlichen Standards.
- Die Mathematik-Aufgaben entsprechen den definierten Grundkompetenzen. Die entwickelten Aufgaben operationalisieren die Deskriptoren der Grundkompetenzen so einfach wie möglich. Als entscheidend erachten sie den Umstand, dass die schweizerischen Bildungsziele in Mathematik im internationalen Vergleich sehr ambitioniert sind und nicht empirisch validiert wurden. Die Grundkompetenzen für die Mathematik verhalten sich empirisch gesehen eher wie Regelstandards.

Der Autor und die Autorin des Auditberichts schlagen die folgenden vier Massnahmen zur Verbesserung und langfristigen Stärkung der ÜGK vor (vgl. Fischbach & Ugen, 2018, 30ff.):

1. *Die Aufgabenentwicklung sollte gestärkt und so die Qualität der Aufgaben verbessert werden.* Die Aufgaben sind, neben den formulierten Mindeststandards, der wichtigste Teil der ÜGK-Erhebungen. Die Ressourcen für die Aufgabenentwicklung sollten deshalb ausgebaut und durch den verstärkten Einbezug von Testexpertinnen und -experten sowie von Lehrpersonen breiter abgestützt werden.

Zudem sollte die Qualität der Testaufgaben durch mehr Pretests und durch valide Übersetzungsprozesse verbessert werden.

2. *Die Grundkompetenzen in Mathematik sollten überdacht werden.* Die vorliegenden Referenzdokumente sind als Mindeststandards sehr ambitioniert und zudem für die Testentwicklung suboptimal formuliert. Auch fehlt eine empirische Validierung der Grundkompetenzen in Mathematik.
3. *Die Projektorganisation sollte überdacht werden.* Bei der ÜGK 2016 war das GS EDK Auftraggeberin und hatte gleichzeitig die operative Leitung inne. Diese Doppelfunktion ist nicht ideal. Die operative Leitung sollte abgegeben und professionalisiert werden.
4. *Das geplante wissenschaftliche Konsortium sollte aufgebaut und eingesetzt werden, ebenso sollte ein unabhängiger wissenschaftlicher Beirat mit (inter-)nationalen Expertinnen und Experten gebildet werden.* Solche Beratungsgremien könnten wesentlich dazu beitragen, die wissenschaftliche Qualität in den verschiedenen Prozessschritten der ÜGK-Erhebungen sicherzustellen sowie die operative Projektleitung politisch und strategisch zu beraten.

Die vom Autor und der Autorin des Auditberichts vorgeschlagenen Massnahmen sind inzwischen wie folgt umgesetzt:

1. Im Bereich der Aufgabenentwicklung sind erste Massnahmen, die den späteren Empfehlungen des Auditberichts entsprechen, bereits im Hinblick auf die ÜGK 2017 umgesetzt worden: Stärkung des Koordinationsmandats der GS ADB; Einbezug von Lehrpersonen und Schülerinnen und Schülern zu einem frühen Zeitpunkt; Verbesserung der Übersetzungsprozesse (allerdings ohne unabhängige doppelte Übersetzungen wie etwa bei PISA) und gründliche Pilotierung der Aufgaben. Im Zusammenhang mit der Aufgabenentwicklung für die ÜGK 2024 im 4. Schuljahr hat der Kosta HarmoS der GS ADB ein Mandat zur Stärkung der Aufgabenentwicklung erteilt. Es umfasst die folgenden Elemente
   - Qualitative Aufgabenerprobungen begleitet von retrospektiven Interviews mit individuellen Schülerinnen und Schülern zur kognitiven Validierung der Aufgaben (Nachvollzug der Lösungsprozesse); dieses Verfahren wird bereits auch bei der Aufgabenentwicklung zur ÜGK 2020 bzw. 2023 umfassend eingesetzt;
   - Begutachtung der Aufgaben durch Lehrpersonen, nicht beteiligte Fachdidaktikerinnen und Fachdidaktiker sowie Heilpädagoginnen und Heilpädagogen aus verschiedenen Sprachregionen;
   - Präpilotierung der Aufgaben mit je sieben Klassen pro Sprachregion in einem standardisierten Verfahren, welches das Verfahren bei der Haupterhebung möglichst weitgehend vorwegnimmt; anschliessende Prozessevaluation und Datenanalyse;
   - Piloterhebung mit sprachregional repräsentativen Schülerstichproben. Die Schwellenwertsetzung wird, wenn immer möglich, aufgrund der Ergebnisse dieser Erhebung stattfinden.

   Der Projektleitung ÜGK stehen die Mittel für die umfangreicheren Piloterierungen zur Verfügung.
2. Parallel zur Aufgabenentwicklung und den Erhebungen werden die freigegebenen Bildungsziele in Mathematik für das 4. Schuljahr verfeinert und konkretisiert. Im Februar 2025 legt das Aufgabenentwicklungsteam einen Schlussbericht dazu vor. Bei nachfolgenden Erhebungen sollen die Bildungsziele der beiden anderen Schulstufen in analoger Weise angepasst werden.
3. Ab Beginn der Planung der ÜGK wurde darauf hingearbeitet, dass eine universitäre Hochschule die wissenschaftliche Projektleitung für die ÜGK-Erhebungen übernimmt. Für die beiden beschlossenen Erhebungen der Jahre 2023 und 2024 (vgl. 1.4) konnte eine entsprechende Kooperationsvereinbarung mit der Universität Bern abgeschlossen werden (vgl. 1.2.3). Das Interfaculty Centre for Educational Research (ICER) nimmt nun die zentralen Koordinationsaufgaben auf der operativen und wissenschaftlichen Ebene wahr. Organe der EDK sind zwar immer noch, aber in wesentlich geringerem Umfang in operative Aufgaben involviert (vgl. dazu Abschnitt 2.4.1).
4. In Abschnitt 1.2.3 wird erläutert, inwiefern der Konzeptentwurf für das WIKO BFE in der aktuellen Struktur aufgenommen wird.

1.3.4 Berichterstattung und Dokumentation

Die Erstberichterstattung zur ÜGK obliegt der EDK. Die Berichte zu den beiden ersten ÜGK-Zyklen werden am 24. Mai 2019 gleichzeitig veröffentlicht (Nidegger, Roos, Petrucci, Verner, Helbling, Arnold, Erzinger,

Pham, Ambrosetti, et al., 2019; Nidegger, Roos, Petrucci, Verner, Helbling, Arnold, Erzinger, Pham, Schönenberger, et al., 2019). Sie enthalten Basisinformationen zur ÜGK, Informationen zum Vorgehen bei der aktuellen ÜGK-Erhebung, zu Lehrplänen und Lehrmitteln sowie, als Kernelement, die Ergebnisse zum Erreichen der Grundkompetenzen und darauf aufbauende Vergleiche zwischen den Kantonen. Darauf folgen sog. adjustierte oder «faire» Vergleiche, bei denen die Effekte der unterschiedlichen sozialen Zusammensetzung der Schülerschaft in den einzelnen Kantonen statistisch neutralisiert (kontrolliert) sind. Der zweite Teil der Berichte enthält für jeden Kanton (bzw. bei zweisprachigen Kantonen für jede Sprachregion) ein «kantonales Porträt», in dem auf wenigen Seiten die wichtigsten Informationen zum Erreichen der Grundkompetenzen pro Fach- bzw. Kompetenzbereich dargestellt sind, zuerst bezogen auf alle Schülerinnen und Schüler, dann unterteilt nach sozialen Gruppen entlang der Merkmale Geschlecht, sozialer Status, zuhause gesprochene Sprache, Migrationsstatus, im 11. Schuljahr auch nach Schulprogramm. Als Teil der Erstberichterstattung veröffentlicht die EDK für ein breiteres Publikum eine Kurzinformation zur ÜGK und zu den beiden Zyklen 2016 und 2017. Sie enthält insbesondere auch Aufgabenbeispiele (EDK, 2019b).

Die Hauptberichte werden von einer Reihe von Berichten zu wissenschaftlichen Aspekten ergänzt, die auf der Website ÜGK Schweiz publiziert sind (Rubrik «Nationale Berichterstattung»): Aufgaben- und Testentwicklung, Stichprobenziehung, Kontextfragebogen (Inhalt, Datenaufbereitung, Analysen).

Gemessen an den Möglichkeiten wird in der Erstberichterstattung sparsam über mögliche Erklärungen für beobachtete Unterschiede bei den Ergebnissen berichtet. Vertiefende Analysen aufgrund der ÜGK-Daten sollen nach Bedarf und Interesse durchgeführt werden. Zu diesem Zweck werden die Datensätze auf FORSbase bzw. SWISSUbase einem wissenschaftlichen Publikum (einschliesslich kantonaler Forschungsstellen) zur Verfügung gestellt. Die PL ÜGK bietet den Kantonen zudem an, Aufträge für vertiefende Analysen zu den ÜGK-Daten anzunehmen.

Das ICER stellt im Anschluss an die Publikation der Ergebnisse interessierten Kantonen auf Wunsch eine erweiterte Dokumentation mit grafischen und tabellarischen Darstellungen der Variablen der Datensätze zur Verfügung. Von diesem Grundangebot haben bisher sechs Kantone profitiert, alle im Zusammenhang mit der ÜGK 2016 (AG, FR, GL, SG, TG, ZH). Insgesamt sieben Kantone haben von Mitgliedern der PL ÜGK vertiefende, auf Forschungsfragen basierende Analysen erstellen lassen, sechs Kantone zur ÜGK 2016 (BL, GE, SG, TI, TG, ZH), vier zur ÜGK 2017 (BL, GL, GE, TI). Im Fall der Kantone Genf und Tessin wurden diese Analysen kantonsintern erstellt.

Die Ergebnisse der ÜGK haben bisher vor allem im Kanton Basel-Landschaft zu entschiedenen Massnahmen geführt. Als Reaktion auf das Abschneiden bei der ÜGK 2016, das als enttäuschend wahrgenommen wurde, hat der Kanton Basel-Landschaft im Jahr 2019 das Programm Zukunft Volksschule lanciert. Dabei handelt es sich um ein umfangreiches Massnahmenpaket, das von 2022 bis 2028 auf der Primarstufe und der Sekundarstufe I umgesetzt werden soll. Besonders im Mittelpunkt steht die Förderung der Sekundarschülerinnen und -schüler im Leistungszug A (Programm mit Grundansprüchen auf der Sekundarstufe I). Sie sollen den Übergang auf die Sekundarstufe II erfolgreich bewältigen können.

Die Autorinnen und Autorinnen der Erstberichterstattung haben verschiedentlich darauf hingewiesen, dass die Ergebnisse der ÜGK auch die Funktion haben, Problemstellen zu lokalisieren, die mit den ÜGK-Daten selbst nicht genügend erforscht werden können, da es sich um Beobachtungsdaten handelt, die in Bezug auf den Nachweis von Kausalitäten naturgemäss limitiert sind. Um kausale Zusammenhänge zu erforschen, müssen Forschungsprojekte lanciert werden, welche die entsprechenden Fragestellungen mit geeigneten wissenschaftlichen Methoden angehen.

1.3.5 Bereitstellung und Nutzung der ÜGK-Daten für Sekundäranalysen

Nach erfolgter Erstberichterstattung im Mai 2019 werden die Daten der ÜGK 2016 und 2017 anonymisiert und in Form eines sog. Scientific-Use-Files (SUF) bei FORS hinterlegt. Die so aufbereiteten Daten lassen keine Rückschlüsse auf die beteiligten Schulen, Klassen und Lehrpersonen bzw. individuellen

Schülerinnen und Schüler zu. Diese Datensätze können insbesondere im Rahmen des nationalen Bildungsmonitorings (SKBF CSRE, 2018) und der kantonalen Qualitätsentwicklungsprozesse genutzt werden. Auf diesem Weg können die Ergebnisse der ÜGK in die kontinuierliche Weiterentwicklung des Bildungssystems einfliessen. Weiter werden sie der Scientific Community für wissenschaftliche Zwecke zur Verfügung gestellt.

Um den Umgang mit den Daten, insbesondere den Datenschutz, die Datensicherheit und die Nutzungsbedingungen vor und nach der Erstberichterstattung zu regeln, beschliesst die Plenarversammlung der EDK am 22. März 2018 ein Datennutzungskonzept. Dieses wird aufgrund der Erfahrungen in der Folge überarbeitet und am 7. Mai 2021 vom Vorstand der EDK verabschiedet (EDK, 2021b). Neben den SUF werden bei FORS z.T. auch geheimzuhaltende, nicht-anonymisierte oder nicht bearbeitete Rohdaten sowie die Testaufgaben gespeichert. In jedem Fall muss eine Datennutzungsanfrage gestellt werden. Anfragen von Forschenden an Hochschulen und anderer staatlich getragener Stellen zur Nutzung des SUF werden in einem vereinfachten Verfahren durch das GS EDK und das ICER beurteilt. Die Nutzung anderer Daten muss der Kosta HarmoS beschliessen. Dies gilt insbesondere auch für Anfragen zur Verknüpfung der ÜGK-Daten mit anderen Datensätzen. Solche Verknüpfungen darf nur das Bundesamt für Statistik BFS ausführen. Vor der Nutzung von ÜGK-Daten muss ein Datennutzungsvertrag unterzeichnet werden, der insbesondere die Kriterien zur Datensicherheit des ÜGK-Reglements festhält.

In den Jahren 2020-2022 ist der Datenbestand bei FORS zu den beiden abgeschlossenen ÜGK-Zyklen weiter ergänzt worden. Insbesondere liegen nun für aufgaben- bzw. itembezogene Forschungsvorhaben auch die Rohantworten der Schülerinnen und Schüler bereit, ebenso alle eingesetzten Aufgaben (freigegebenen und unter Verschluss gehaltene). Zudem können die beiden SUF nun in zwei Fassungen genutzt werden: als Datensatz mit fehlenden Daten (z.B. aufgrund von Antwortverweigerungen) und als Sets von jeweils 20 statistisch imputierten vollständigen Datensätzen.

Im Winter 2021-2022 hat die FHGR ein Archiv zu den beiden abgeschlossenen ÜGK-Zyklen erstellt, in dem die digitalen Aufgaben in lauffähigem Zustand archiviert werden. Hinzu kommen die Log-Dateien der Testsoftware sowie die elektronischen Fragebogen mit den Metadaten. Das Archiv wird zudem mit Blick auf die Nutzung dokumentiert. Für die Nutzung dieser Daten zu Forschungszwecken können ebenfalls Anfragen gestellt werden. Die Archivierung dieser Daten ist vorerst bis Ende 2025 vertraglich vereinbart.

Die Dokumentation der Daten bei FORS ist zwar umfassend, für Forschende, die nicht selbst an den ÜGK-Zyklen beteiligt waren, aber eher unübersichtlich. Mitglieder der PL ÜGK haben bei der Datennutzung verschiedentlich Unterstützung geleistet.

FORS sammelt die Datennutzungsanfragen mit kurzen Beschreibungen der Vorhaben, führt aber nicht Buch über den tatsächlichen Bezug und die tatsächliche Verwendung der Daten. Relativ viele Anfragen stammen von Forschenden, die einen Bezug zur ÜGK haben. Im Heft 2/2020 des *Swiss Journal of Educational Research* sind beispielsweise sechs Beiträge von Forschenden aus dem Umfeld der ÜGK erschienen, die anhand von Daten der ÜGK 2016 verschiedene Zusammenhänge zwischen den gemessenen Kompetenzen und Bedingungsfaktoren von Leistungen untersuchen: schulischer Kontext, Unterrichtsqualität, Noten, Motivation, Wohlbefinden und Selbsteinschätzung der Fähigkeiten. Eine Reihe von Anfragen für die Daten der ÜGK 2016 stammt aus dem Umkreis des Projekts TREE (*Transition from Education to Employment*), mit dem es eine Zusammenarbeit bei der Erstellung des Kontextfragebogens gab. Die Schweizerische Koordinationsstelle für Bildungsforschung SKBF verwendet die Daten beider ÜGK-Zyklen in Kombination mit anderen Datenbeständen bei Auswertungen für den *Bildungsbericht Schweiz 2023*. Einzelne weitere Forschungsvorhaben haben eine soziologische Ausrichtung oder gehören in den Berufsbildungsbereich. Drei Vorhaben nutzen die Daten der ÜGK 2016 bzw. 2017 für Studien zu den Testaufgaben selbst. Verschiedene weitere Datennutzungsanfragen stehen im Zusammenhang mit Qualifikationsarbeiten an Hochschulen.

1.4 Beschlossene zukünftige ÜGK-Erhebungen

Am 22. Juni 2017 beschliesst die EDK-Plenarversammlung, in den Jahren 2020 und 2022 zwei weitere ÜGK-Erhebungen durchzuführen (EDK, 2017). Die zu testenden Fachbereiche legt sie in ihren Plenarversammlungen vom 26. Oktober 2018 bzw. 25. Oktober 2019 fest (EDK, 2018b, 2019c). Im Jahr 2020 soll am Ende des 11. Schuljahres das Erreichen der Grundkompetenzen in der Schulsprache sowie der ersten und zweiten Fremdsprache überprüft werden. Damit soll dieselbe Kohorte nach 2017 zum zweiten Mal in der Schulsprache und ebenfalls zum zweiten Mal in der ersten Fremdsprache getestet werden. Für die ÜGK 2022 ist vorgesehen, erstmals im 4. Schuljahr das Erreichen der Grundkompetenzen zu überprüfen, und zwar in den Fachbereichen Schulsprache und Mathematik. Wie in Abschnitt 1.2.3 ausgeführt, ist bei diesen beiden ÜGK-Zyklen erstmals die neue Projektstruktur mit der Universität Bern bzw. dem ICER an der Spitze mit der Durchführung betraut.

Aufgrund der Corona-Pandemie beschliesst die EDK-Plenarversammlung am 1. April 2020, auf die unmittelbar bevorstehende Haupterhebung zur ÜGK 2020 zu verzichten (EDK, 2020a). Am 25. Juni 2020 beschliesst die Plenarversammlung einen neuen Zeitplan, der auch mit der ebenfalls verschobenen PISA-Erhebung im Jahr 2022 koordiniert ist (EDK, 2020b). Die ÜGK 2020 wird auf das Jahr 2023 und die ÜGK 2022 auf das 2024 verschoben.

Im Folgenden werden, soweit bekannt, die Eckwerte der beiden beschlossenen ÜGK-Erhebungen beschrieben.

1.4.1 ÜGK 2023: Schulsprache sowie erste und zweite Fremdsprache im 11. Schuljahr

Aufgaben- und Testentwicklung

Die Tests im Rahmen der ÜGK 2023 im 11. Schuljahr betreffen die Schulsprache sowie die erste und zweite Fremdsprache. Als Schulsprache wird je nach Region Deutsch, Französisch oder Italienisch getestet, und zwar anhand der Kompetenzbereiche Lesen und Orthografie. Bei den Fremdsprachen werden bei jeweils einem Sprachenpaar die Kompetenzen im Hör- und Leseverstehen überprüft. Bei diesen Sprachenpaaren handelt es sich in der Deutschschweiz um Französisch und Englisch, in der Westschweiz um Deutsch und Englisch und im Tessin um Französisch und Deutsch.

Im Fachbereich Schulsprache entwickelt die GS ADB die Testaufgaben in Zusammenarbeit mit derselben Gruppe von Fachdidaktikerinnen und Fachdidaktikern an pädagogischen Hochschulen (PH FHNW, HEP Vaud, SUPSI) sowie Mitarbeitenden des Dipartimento dell'educazione, della cultura e dello sport (DECS) des Kantons Tessin wie bei der ÜGK 2017, sodass auf Vorarbeiten und Vorkenntnisse zurückgegiffen werden kann. Bei den Fremdsprachen werden mehr Institutionen einbezogen als im Jahr 2017, um die Aufgabenentwicklung breiter abzustützen. Die GS ADB arbeitet neu mit Fachdidaktikerinnen und Fachdidaktikern sowie Testspezialistinnen und Testspezialisten der folgenden Institutionen zusammen: Institut für Mehrsprachigkeit der Universität Freiburg und der PH Freiburg, PH FHNW, PH Luzern und PH Zürich. Die Aufgabenentwicklung wird zudem von einer Expertengruppe von Fachdidaktikerinnen und Fachdidaktikern aus der französischen und italienischen Schweiz begleitet.

Um frühe Hinweise auf die Qualität der Aufgaben zu erhalten, werden alle Aufgaben mit einer Reihe von einzelnen Schülerinnen und Schülern qualitativ erprobt. Dabei werden zum Aufgabenlösungsprozess retrospektive Interviews geführt, um Einsicht in die Lösungsprozesse zu erhalten (kognitive Validierung). Im Frühjahr 2019 werden die Testaufgaben schliesslich an einer nationalen Stichprobe von über 2'000 Schülerinnen und Schülern erprobt. Aufgrund von psychometrischen Analysen der Daten werden einzelne Aufgaben angepasst, bevor sie für die Haupterhebung bereitgestellt werden.

Insgesamt können im Fachbereich Schulsprache Aufgaben im folgenden Umfang eingesetzt werden: Kompetenzbereich Lesen: 44 Items; Kompetenzbereich Orthografie (sprachspezifische Tests): Deutsch 50, Französisch 54 und Italienisch 49 Items.

In den Fremdsprachen kommen im Kompetenzbereich Leseverstehen 47 Items und beim Hörverstehen 52 Items zum Einsatz. Diese Items liegen entsprechend den getesteten Fremdsprachen in den folgenden Sprachversionen vor: geteste Sprache (Zielsprache) Deutsch, Anweisungen und Aufgabenstellungen auf Französisch oder Italienisch; Zielsprache Französisch, Anweisungen und Aufgabenstellungen auf Deutsch oder Italienisch; Zielsprache Englisch, Anweisungen und Aufgabenstellungen auf Deutsch oder Französisch. Da sich die in den Grundkompetenzen formulierten Anforderungen zwischen der ersten und der zweiten Fremdsprache nicht unterscheiden, werden jeweils in allen Zielsprachen die gleichen Testaufgaben eingesetzt. Die Wahl der Texte für die Testaufgaben ist das Ergebnis einer umfangreichen Analyse der Lehrwerke für die drei Zielsprachen.

Kontextualisierung

Die Kontextualisierung der ÜGK 2020 besteht wiederum aus einem Kontextfragebogen für Schülerinnen und Schüler. Inhaltlich werden nach einer gründlichen Überprüfung viele Konstrukte und Variablen aus dem Fragebogen zur ÜGK 2017 übernommen. Die PL ÜGK arbeitet darauf hin, (möglichst) in allen ÜGK-Erhebungen ein festes «Basismodul» an Fragebogenelementen umzusetzen, auch um ÜGK-übergreifende Analysen zu ermöglichen. Mit dem Fragebogen werden u.a. Informationen zur sozialen Herkunft, zur Familien- und zur Schulsituation der Schülerinnen und Schüler erhoben. Darüber hinaus gibt es Fragen zu den individuellen Lernvoraussetzungen und -merkmalen, zur Motivation, zum eigenen Interesse sowie zu verschiedenen Merkmalen des Unterrichts (z.B. Unterrichtsaktivitäten).

Durchführung

Die ÜGK-Erhebung im Frühjahr 2023 soll wiederum nach einem standardisierten Vorgehen durch geschulte Testverantwortliche durchgeführt werden. Wie bereits in der ÜGK 2017 werden die Schülerinnen und Schüler während der ÜGK-Erhebung auf mitgebrachten Tablets arbeiten. Die individuellen Tests bestehen aus drei Blöcken à 40 Minuten (Schulsprache, erste bzw. zweite Fremdsprache). Im Anschluss daran füllen die Schülerinnen und Schüler den Kontextfragebogen aus.

An der geplanten ÜGK 2023 nehmen, abgesehen vom Kanton ZG, alle Kantone der Schweiz teil. Aus den insgesamt rund 80'000 Schülerinnen und Schülern, die das 11. Schuljahr in den teilnehmenden Kantonen besuchen, werden für die ÜGK-Erhebung kantonal repräsentative Stichproben im Umfang der ersten ÜGK-Zyklen gezogen (mind. 20’000 Personen). Schülerinnen und Schüler, die im 11. Schuljahr keinen Fremdsprachenunterricht mehr besuchen, werden grundsätzlich nicht von einer Teilnahme an der ÜGK-Erhebung dispensiert.

1.4.2 ÜGK 2024: Schulsprache und Mathematik im 4. Schuljahr

Aufgaben- und Testentwicklung

Die Aufgabenentwicklung im Hinblick auf die ÜGK 2024 schliesst an die im Rahmen der Machbarkeitsstudie zur Überprüfung der Grundkompetenzen im 4. Schuljahr (Angelone et al., 2019) geleisteten Vorarbeiten an. Die Machbarkeitsstudie wurde 2018 vom GS EDK in Auftrag gegeben, um als Grundlage für die weitere Planung die Möglichkeiten und Herausforderungen aufzuzeigen, die sich bei einer standardisierten und tabletbasierten Überprüfung der Grundkompetenzen bei Kindern im Alter von ca. acht Jahren ergeben. Im Rahmen dieser Machbarkeitsstudie haben Fachdidaktikerinnen und Fachdidaktiker erste theoretische Test-Frameworks und Beispielaufgaben für die Fachbereiche Mathematik, Naturwissenschaften und Schulsprache erarbeitet. Sie zeigen, welche der formulierten Grundkompetenzen im Rahmen einer ÜGK im 4. Schuljahr getestet werden können. Zudem hat die GS ADB in vier Schulklassen eine standardisierte Testdurchführung und die PH St. Gallen den Einsatz eines tabletbasierten Schülerfragebogens erprobt.

In Zusammenarbeit mit der GS ADB erarbeiten Fachdidaktikerinnen und Fachdidaktiker der beiden Fachbereiche in den Jahren 2020 und 2021 die Testaufgaben. Bei der Schulsprache kooperieren die PH FHNW, die HEP Vaud, die PH Zug, die SUPSI und das DECS des Kantons Tessin. Bei der Mathematik

beteiligen sich das IRDP, die PH Thurgau, und die SUPSI. Diese wiederum arbeiten mit Expertinnen und Experten weiterer Institutionen zusammen (Universität Genf, HEP Vaud, PHLU, PHZH).

Wie bereits in Abschnitt 1.3.3 unter den Folgerungen aus dem Auditbericht zur ÜGK 2016 dargestellt, wird die Aufgabenentwicklung zur ÜGK 2024 durch ein Mandat der EDK wesentlich gestärkt, das der Kosta HarmoS am 13. November 2020 zu Handen der ADB verabschiedet hat. Es umfasst neben Massnahmen zur Verbesserung der Aufgabenqualität auch die Verfeinerung und Konkretisierung der existierenden Beschreibungen der Bildungsziele für die Mathematik in Form eines Test-Frameworks mit Beispielaufgaben. Dieses Framework bildet bei der ÜGK 2024 ein Zwischenglied zwischen den freigegebenen Bildungsgzielen und der Aufgaben- und Testentwicklung.

Kontextualisierung

Im Zusammenhang mit der Kontextualisierung der ÜGK im 4. Schuljahr erprobt die PL ÜGK neue Möglichkeiten, um trotz des jungen Alters der Schülerinnen und Schüler genügend zuverlässige Kontextdaten erheben zu können. Zum einen wird unter Einbezug von Kindern der Zielgruppe eine kindgerechte Form für den Fragebogen gesucht, der ausgefüllt werden soll. Zum Beispiel führt eine sprechende Eule durch den Fragebogen. Zum anderen versucht man ergänzende Informationsquellen zu erschliessen. In diesem Zusammenhang wird erforscht, inwiefern Registerdaten des BFS genutzt werden können. Unter Einsatz von eigenen Projektmitteln nutzt das ICER zudem die Piloterhebung zur ÜGK 2024, die im Frühjahr 2022 stattfindet, um Erkenntnisse zum Einsatz und zum Nutzen eines Eltern- und eines Schulleitungsfragebogens zu gewinnen. Mit belastbaren Ergebnissen zu diesen Bemühungen, insbesondere auch zu Rücklaufquoten der Fragebogen, ist im Verlauf des Jahres 2022 zu rechnen.

Durchführung

Sofern die Piloterhebung im Frühjahr 2022 erfolgreich verläuft, soll die ÜGK 2024 im selben Rahmen durchgeführt werden wie zuvor die ÜGK 2017 und 2023.

1.4.3 Finanzierung der ÜGK und der ADB

Die ÜGK wird seit 2014 durch jährliche Kantonsbeiträge in nahezu identischer Höhe finanziert. Weil wegen der Corona-Pandemie Erhebungen verschoben und Arbeiten wiederholt werden mussten, beschloss die EDK-Plenarversammlung am 24. Juni 2021 eine Erhöhung der Kantonsbeiträge, die für die Jahre 2022-25 vorgesehen sind.

Bisherige und geplante Finanzierung der ÜGK von 2014-2025 (vier ÜGK-Erhebungen):

	Beitrag pro Jahr	Gesamtbetrag
Kantonsbeiträge 2014-16 (3 Jahre)	CHF 1‘123‘333	CHF 3‘370‘000
Kantonsbeiträge 2017-25 (9 Jahre)	CHF 1‘130‘000	CHF 10‘170‘000
Corona-bedingte Erhöhung der Kantonsbeiträge 2022-25	CHF 232'500	CHF 930‘000
Total Kantonsbeiträge 2014-25 ÜGK		**CHF 14‘470‘000**

Wie in Abschnitt 1.2.2 ausgeführt, übernimmt die Aufgabendatenbank wichtige Aufgaben für die ÜGK. Sie hat allerdings gemäss Reglement eine Reihe von Aufgaben, die nicht mit der ÜGK zu tun haben und wird getrennt von der ÜGK finanziert. Der ADB wurden bisher für die Jahre 2012-25 die folgenden Beträge zur Verfügung gestellt:

	Beitrag pro Jahr	Gesamtbetrag

Aufbaukredit aus Mitteln des Bildungsmonitorings (2012-13)		CHF 765‘000
Kantonsbeiträge 2014-25 (12 Jahre)	CHF 370‘000	CHF 4‘440‘000
Total Aufbaukredit und Kantonsbeiträge 2014-25 ADB		**CHF 5‘205‘000**

Mit den Beiträgen der Jahre 2020 bis 2025 an das ÜGK-Projekt und an die Aufgabendatenbank wird die Durchführung der ÜGK-Zyklen 2023 und 2024 (ursprünglich 2020 und 2022) sichergestellt.

Insgesamt sind die Ausgaben der Schweiz für das Bildungsmonitoring im Vergleich zu den Ausgaben der übrigen D-A-CH-Länder gering.

2 Analyse und Diskussion

Im folgenden Kapitel stützt sich der Kosta HarmoS insbesondere auf ein wissenschaftliches Gutachten, welches das Generalsekretariat der EDK bei Expertinnen und Experten aus den Bereichen der Governance im Bildungsbereich sowie der Large-Scale-Assessments in Auftrag gegeben hat. Die Gutachtengruppe setzte sich wie folgt zusammen:

- Prof. Dr. Katharina Maag Merki, Universität Zürich (Koordination)
- Dr. Anja Giudici, Universität Oxford
- Prof. i.R. Dr. Eckhard Klieme, DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation, Frankfurt/Main
- Prof. Dr. Hans Anand Pant, Humboldt-Universität zu Berlin
- Prof. Dr. Claudia Schreiner, Universität Innsbruck

Kai Schudel, wissenschaftlicher Mitarbeiter der Universität Zürich, hat die Gruppe bei ihrer Arbeit unterstützt.

Als Grundlagen für das folgende Kapitel dienen zudem Stellungnahmen von weiteren externen Expertinnen und Experten, namentlich von Olivier Maradan, früherer Leiter des Koordinationsbereichs Obligatorische Schule im Generalsekretariat der EDK; Christian Nidegger, u.a. früherer Nationaler Projektmanager der PISA-Studie und Hauptautor zur Berichterstattung zu den ÜGK 2016 und 2017, sowie Stefan Wolter, Direktor der Schweizerischen Koordinationsstelle für Bildungsforschung SKBF und verantwortlicher Autor des «Bildungsbericht Schweiz» (SKBF CSRE, 2018), dem zentralen Instrument der Bildungsberichterstattung. Zudem sind zahlreiche Hinweise des ICER bzw. der Projektleitung ÜGK eingeflossen.

2.1 Ansatz, Ziele und Konzepte der ÜGK

In den folgenden Abschnitten werden grundlegende Merkmale der ÜGK kritisch diskutiert: der Ansatz der Bildungssteuerung über den Output; die hauptsächlichen Ziele und Funktionen; Interpretation und Umgang mit den Ergebnissen; der Fokus auf den Grundkompetenzen sowie die Beschränkung auf vier Fachbereiche.

2.1.1 Systemsteuerung über den Output

Mit der ÜGK sollen die Qualitätssicherung und die inhaltliche Harmonisierung der kantonalen Bildungssysteme auf nationaler Ebene über gemeinsame nationale Bildungsstandards erfolgen, die den Charakter von Leistungsstandards haben (vgl. Abschnitt 1.1).

Der Output der Schulsysteme wird anhand der Leistungen der Schülerinnen und Schüler gemessen und verglichen. Damit reiht sich die ÜGK in vergleichbare Systeme der Outputsteuerung in verschiedenen europäischen Ländern ein, die seit den 2000er Jahren etabliert wurden.

Das Gutachten setzt sich kritisch mit der Frage auseinander, ob dieser Ansatz noch zeitgemäss ist. Es kommt zum Schluss, dass die Einrichtung entsprechender Bildungsmonitorings anspruchsvoll ist und nicht ohne Weiteres zu einem Bildungssystem führt, in dem alle Schülerinnen und Schüler die gesetzten Ziele erreichen. Die Gutachtengruppe spricht sich aber klar für die Beibehaltung der ÜGK aus unter der Voraussetzung, dass substanzielle Anpassungen am Gesamtdesign, an der institutionellen Verankerung und Gestaltung sowie bei der Zuweisung von finanziellen und personellen Ressourcen vorgenommen werden (siehe folgende Abschnitte).

Für den Kosta HarmoS ist die Steuerung gewisser Aspekte der kantonalen Schulsysteme durch Leistungsmessungen mit Blick auf den verfassungsmässigen Harmonisierungsauftrag unabdingbar. Ein Verzicht auf Leistungsmessungen würde die Harmonisierung gefährden und die gesamtschweizerische Zielsetzung der Qualitätsentwicklung der obligatorischen Schule in Frage stellen. Soll die Leistungsmessung aber ihre Wirkung in der Systemsteuerung entfalten, so muss sie im Sinne der im

Gutachten enthaltenen Vorschläge, die von den übrigen befragten Experten unterstützt und zum Teil weiter akzentuiert wurden, überarbeitet werden.

2.1.2 Zentrale Funktionen der ÜGK

Wie in Abschnitt 1.2.1 ausgeführt, besteht das Ziel der ÜGK gemäss ÜGK-Reglement darin, die Leistungen der obligatorischen Schule zu evaluieren und den Kantonen auf der Systemebene Informationen zum Grad des Erreichens der Bildungsziele zu liefern. Mit der Erhebung von Kontextdaten können Zusammenhänge zwischen Leistungsunterschieden und Kontextmerkmalen identifiziert werden.

Die Gutachtengruppe hat bei den Befragungen von bildungspolitischen Akteurinnen und Akteuren festgestellt, dass von der ÜGK folgende drei Funktionen erwartet werden:

1. Systemevaluation
2. Bereitstellung von Erklärungsansätzen für Leistungsunterschiede
3. Bereitstellung von Handlungswissen für die schulische Praxis

Die *Funktion der Systemevaluation* ist weitgehend unbestritten, wenn auch zum Teil erhebliche Vorbehalte in Bezug auf die bisherige Umsetzung festgestellt wurden.

Die *Bereitstellung von Erklärungsansätzen für Leistungsunterschiede* zwischen Regionen, Kantonen und sozialen Gruppen innerhalb der Schülerschaft ist aus Sicht der Gutachtengruppe ebenso wie der weiteren befragten Experten zentral, wird aber mit der heutigen Umsetzung nur bedingt erreicht. Soll die ÜGK effektiv zu einer Qualitätsentwicklung beitragen können, so muss die Erklärungsfunktion gestärkt werden. Dies würde neben dem realen auch den wahrgenommenen Nutzen der ÜGK und damit deren Akzeptanz steigern. Die dafür vorgeschlagenen Massnahmen werden in den nachfolgenden Abschnitten vorgestellt.

Für den Kosta HarmoS ist zentral, dass Leistungsmessungen qualitativ hochstehend und glaubwürdig sind und dass ihr Nutzen für die Systemevaluation, das Bildungsmonitoring und die Bildungssteuerung gesteigert wird. Sie sollen Grundlage für eine evidenzbasierte Politik sein.

Die *Bereitstellung von Handlungswissen für die schulische Praxis,* die gemäss der Befragung ebenfalls von der ÜGK erwartet wird, unterstützt das Gutachten ebenso wenig wie die Experten oder der Kosta HarmoS. Eine Untersuchung wie die ÜGK, die nicht auf die Individualebene zielt, kann dies nicht leisten. Sie müsste anders konzipiert werden und Aufgaben übernehmen, die gemäss HarmoS-Konkordat den Regionen und Kantonen zufallen.

2.1.3 Interpretation der Ergebnisse

Die Bundesverfassung verpflichtet die Kantone zur Harmonisierung der Ziele der Bildungsstufen. Mit der Überprüfung des Erreichens dieser Ziele soll der Verfassungsauftrag umgesetzt werden.

Das Gutachten weist darauf hin, dass im Zusammenhang mit den Ergebnissen der bisherigen ÜGK-Zyklen nicht festgelegt worden ist, bei welchem Erreichungsgrad die Grundkompetenzen in genügendem Mass erreicht sind und bis zu welchen Unterschieden zwischen den Kantonen die kantonalen Bildungssysteme als harmonisiert gelten können. Es schlägt deshab vor, für diese Beurteilung einen Mechanismus vorzusehen.

Vor dem Hintergrund, dass Grundkompetenzen definitionsgemäss von (praktisch) allen Schülerinnen und Schülern erwartet werden, obliegt es aus Sicht des Kosta Harmos der Bildungspolitik, bei Vorliegen der Ergebnisse einer Überprüfung zu entscheiden,

- ob aufgrund des Anteils an Schülerinnen und Schülern, welche die Mindestkompetenzen nicht erreichen, Massnahmen ergriffen werden müssen, und wenn ja, durch welche Ebene(n);
- inwieweit der verfassungsmässige Harmonisierungsauftrag erfüllt ist.

2.1.4 Fokus auf den Grundkompetenzen und Blick auf das gesamte Kompetenzspektrum

Die ÜGK erlauben aktuell aufgrund des Fokus auf den Grundkompetenzen eine Aussage über den Erreichungsgrad der Grundkompetenzen, aber wenig bis keine Aussagen zu den Leistungen der Gesamtheit der Schülerinnen und Schüler, sofern die Tests tatsächlich, wie verlangt, auf die Grundkompetenzen fokussieren. Ein klarer Fokus auf dem unteren Fähigkeitsbereich war jedoch in den bisherigen Tests eher die Ausnahme, sodass die Tests zu mehreren Kompetenzbereichen unbeabsichtigt auch zwischen den stärkeren Schülerinnen und Schülern differenzierten.

Seit den Anfängen der ÜGK wird von verschiedener Seite angeregt, mit der ÜGK nicht nur das Erreichen der Grundkompetenzen zu überprüfen, sondern Tests einzusetzen, die das ganze Kompetenzspektrum zuverlässig erfassen. Dies würde bedeuten, dass das ganze Kompetenzspektrum für die Testkonstruktion und -interpretation auch beschrieben werden müsste, dies vorzugsweise durch Kompetenzskalen mit mehreren Niveaus, ähnlich den PISA-Skalen. Die Akteurinnen und Akteure, die für das Gutachten befragt wurden, sind in der Frage, ob der Fokus der ÜGK auf den Grundkompetenzen (d.h. auf einer Kompetenzschwelle) oder auf dem ganzen Kompetenzspektrum liegen soll, gespalten.

Das Gutachten empfiehlt, bei der Überprüfung der Grundkompetenzen zu bleiben und in der Erstberichterstattung weiterhin nur zu den Anteilen der Schülerinnen und Schüler zu kommunizieren, welche die Grundkompetenzen erreicht haben. Dies mit folgenden Begründungen:

- Das Vorgehen weist eine breite Akzeptanz auf.
- Die ÜGK können auf diese Weise dazu beitragen, dass praktisch alle Schülerinnen und Schüler tatsächlich Mindestkompetenzen erreichen.
- Bei einem Wechsel zu Kompetenzskalen müssten die Bildungsziele neu festgelegt werden, was in Anbetracht der Unterschiede zwischen den sprachregionalen Lehrplänen vermutlich schwierig wäre.
- Die bisherigen ÜGK-Tests weisen eine bereits relativ starke Streuung auf, die Sekundäranalysen in einem breiteren Leistungsspektrum erlaubt.
- Bei der Skalierung der Testergebnisse werden bereits heute kontinuierliche Testskalen erstellt, die auch für Sekundäranalysen zur Verfügung gestellt werden.

Auch der Kosta HarmoS ist der Auffassung, dass eine Überprüfung des Erreichens der Grundkompetenzen im Hinblick auf die Gewährleistung der Grundkompetenzen der Schülerinnen und Schüler das zentrale Ziel der ÜGK bleiben soll. Gleichzeitig kommt er als Steuergremium der ÜGK zum Schluss, dass aus folgenden Gründen künftig mit den Erhebungen mehr für das Bildungsmonitoring und die evidenzbasierte Bildungssteuerung nützliche Daten generiert werden müssen:

- Das Verhältnis zwischen den für die Erhebungen aufgewendeten Ressourcen und der Nutzbarkeit der Daten ist unbefriedigend, wenn die Tests ausschliesslich auf die Grundkompetenzen ausgerichtet sind.
- Das resultierende Steuerungswissen beschränkt sich auf eine zu kleine Gruppe.
- Die mit den Erhebungen verbundenen politischen und öffentlichen Erwartungen können nicht eingelöst werden: Die ÜGK wurde auch als Ersatz für die früher üblichen Kantonsvergleiche im Rahmen der PISA-Studie verstanden (vgl. EDK, 2009, 2012a). PISA zeigt aufgrund der mehrstufigen Niveauskalen die Verteilung der Schülerinnen und Schüler auf die Niveaus und analysiert auch die Streuungsbreite kritisch.

ÜGK-Tests sollen künftig Kompetenzen über das ganze Kompetenzspektrum hinweg zuverlässig erfassen und müssen auch zwischen Schülerinnen und Schülern im oberen und untersten Bereich der Skala differenzieren können. Da einer Erhebung, die Kompetenzen über das ganze Leistungsspektrum testet, eine entsprechende Skalenbeschreibung zugrunde liegen muss, sollten die geltenden Bildungsziele für die Zwecke der Testkonstruktion und der Interpretation der Ergebnisse ergänzt werden (vgl. für Weiteres Anhang 1: Weitere Aspekte der Umsetzung der ÜGK, Abschnitt *Konkrete Grundlagen für die Testentwicklung*). Diese ergänzenden Skalenbeschreibungen müssen aus Analysen der Lehrpläne und eventuell Lehrmitteln der verschiedenen Sprachregionen abgeleitet werden, damit allseits faire

Erwartungen an die Kompetenzen formuliert werden. Die 2011 von der EDK freigegebenen Bildungsziele selbst müssen dabei nicht angepasst werden. Sie behalten für Aussagen zur Erreichung der Grundkompetenzen und damit die bildungspolitische Zielsetzung der Harmonisierung zumindest in Bezug auf Grundkompetenzen und als Grundlage für die sprachregionalen Lehrpläne ihre Gültigkeit.

Die im Gutachten erwähnte Tatsache, dass auch bisherige ÜGK-Tests eine breite Streuung der Schülerleistungen zeigen, ist aus Sicht des Kosta HarmoS nicht relevant, weil es sich um einen aus der Sicht der Testentwicklung unerwünschten Effekt handelt. Die betreffenden Tests fokussierten weniger als erwünscht auf die Schülerinnen und Schüler im unteren Teil des Leistungsspektrums. Grundsätzlich gilt, dass die Ansprüche der Testaufgaben dem Testziel entsprechen müssen. Wenn konsequent nur das Erreichen von realistischerweise zu erwartenden Grundkompetenzen getestet würde, könnte der Test die Leistungen der stärkeren Schülerinnen und Schüler nicht differenziert erfassen. Mittlere, gute und sehr gute Schülerinnen und Schüler könnten alle (fast) perfekte Ergebnisse erreichen. Die resultierenden Testskalen eignen sich nur ansatzweise, um Zusammenhänge mit anderen Variablen, zum Beispiel dem sozialen Hintergrund, aufzuzeigen oder Verschiebungen ausgewählter Gruppen innerhalb der Kohorte von einem Testzeitpunkt zum nächsten (z.B. zwischen dem 4. und dem 8. Schuljahr) zu analysieren. Solche Analysen stehen aber im Zentrum des Bildungsmonitorings und der quantitativ orientierten Bildungsforschung.

2.1.5 Beschränkung auf die Grundkompetenzen in den vier Fachbereichen

Das Gutachten untersuchte die Akzeptanz der Beschreibungen der im Jahr 2011 freigegebenen Bildungsziele (Grundkompetenzen) und beurteilte diese als insgesamt positiv. Die Gutachtengruppe empfiehlt aber, die 2021 für das 4. Schuljahr begonnene Überarbeitung der Beschreibung der Grundkompetenzen in Mathematik auch für die übrigen Stufen fortzusetzen. Der Kosta HarmoS unterstützt diesen Vorschlag. Die im Fachbereich Fremdsprachen erwarteten Grundkompetenzen wurden zu einem Zeitpunkt festgelegt, als noch nicht zuverlässig abgeschätzt werden konnte, wie sich die Umsetzung der Sprachenstrategie, insbesondere der Unterricht von zwei Fremdsprachen bereits auf Primarstufe, auf die Kompetenzen auswirken würde. Die Ergebnisse der ÜGK-Zyklen 2017 und 2023 sollten daraufhin ausgewertet werden, ob die Bildungsziele tatsächlich Grundkompetetenzen beschreiben.

Das Gutachten befasst sich auch mit der Frage, ob die Fachbereiche, in denen getestet wird, richtig gewählt sind und ob neben Schulsprache, Mathematik, Fremdsprachen und Naturwissenschaften auch für weitere Fachbereiche Bildungsziele formuliert werden sollten, wie dies als Option im ÜGK-Konzept vorgesehen ist (vgl. 1.1.2, 1.1.4).

In der Befragung zeigt sich eine sehr hohe Zustimmung für die Fachbereiche Schulsprache und Mathematik. Etwas tiefer liegt die Zustimmung bei den Fremdsprachen, und einen zusätzlichen Abstand weisen die Naturwissenschaften auf. Dagegen werden kaum Vorschläge für weitere Fachbereiche gemacht, zu denen nationale Bildungsziele erarbeitet werden sollten. Die Gutachtengruppe sieht denn auch keinen Anlass für eine Ausweitung auf weitere Fachbereiche. Sie weist dagegen darauf hin, dass bei den Naturwissenschaften das Funktionieren der Tests erst noch zu zeigen ist, vor allem wenn, wie dies in der Validierungsstudie von 2007 der Fall war, in der grossen ÜGK-Stichprobe auch Experimentieraufgaben eingesetzt werden sollen.

In der Bildungsforschung werden die Kompetenzen in Schulsprache und Mathematik als hinreichende Prädiktoren für den späteren Bildungserfolg angesehen. Falls die Kapazitäten im ÜGK-Projekt für ein regelmässiges Testen (im Sinne von 2.1.1) von mehr Fachbereichen nicht ausreichen sollten, wären aus der Sicht des Direktors der SKBF die Prioritäten bei diesen Fachbereichen zu setzen. Aus der Sicht der Schule lässt sich eine solche Prioritätensetzung zudem mit der Promotionsrelevanz begründen.

Der Kosta HarmoS erkennt aus bildungspolitischer Sicht keinen Anlass für eine Veränderung des Katalogs an Fachbereichen, die im Rahmen der ÜGK überprüft werden sollen. Der Stellenwert der Fremdsprachen wird als hoch eingeschätzt. Bei der ÜGK 2023 ist es möglich, die Schulsprache und zwei Fremdsprachen im selben Zyklus zu testen. Es ist erstrebenswert, dass Kombinationen von Fachbereichen auch in Zukunft

umgesetzt werden können. Die Gutachtengruppe schlägt vor, Schulsprache und Fremdsprachen (wie bisher) bzw. Mathematik und Naturwissenschaften jeweils in denselben Erhebungen zu testen, wobei die Machbarkeit von ÜGK-Tests für die Naturwissenschaften noch zu zeigen sei.

2.2 Erhebungsdesign der ÜGK

Die bisherigen und bis 2024 geplanten ÜGK-Zyklen haben den Charakter von initialen Einzelmessungen (Querschnittstudien). Für die Sprachen war (vorerst) einmalig ein Kohorten-Längsschnittdesign geplant, das aufgrund der Pandemie nicht umgesetzt werden konnte. Für Trendmessungen analog zur PISA-Studie[6] wurde bisher keine feste Kadenz festgelegt.

2.2.1 Langfristplanung zur Gewährleistung von Trendmessungen

Das Gutachten bemängelt das Fehlen einer verlässlichen Planung der ÜGK und weist auf den grossen bildungspolitischen Nutzen eines langfristigen Programms hin:

- Es generiert Daten zu Entwicklungen im Bildungssystem und macht Trendaussagen möglich.
- Es macht die Ergebnisse für die Bildungspolitik und -verwaltung, die wissenschaftlichen Partner, das Bildungsmonitoring sowie für die weitere Forschung systematisch nutzbar und deshalb attraktiv.
- Es erlaubt eine Koordination mit den regionalen und kantonalen Monitoringsystemen.
- Es bringt mehr Kontinuität und vermutlich bessere Qualität in die Entwicklung von Aufgaben, Tests und Kontextfragebogen.
- Es sichert Kontinuität bei den Fachexpertinnen und Fachexperten, welche die ÜGK umsetzen.
- Es macht die ÜGK im öffentlichen Diskurs relevant.

Trendstudien bieten im Vergleich zu *ad hoc* beschlossenen Querschnittstudien für die Bildungssteuerung wichtige Vorteile. Ohne die Einbindung in ein Trenddesign kann aus Monitoring-Erhebungen viel weniger Steuerungswissen gewonnen werden. Kantonsvergleiche sowie Vergleiche unter Einbezug von Kontextmerkmalen aufgrund einzelner Erhebungen lassen Ursachen und Ansatzpunkte für Massnahmen höchstens vermuten sowie die Effekte von Massnahmen nur ansatzweise verfolgen. Trendstudien sind für die weiterführende Ursachenforschung ein entscheidender Ausgangspunkt. Sie zeigen, ob sich über die Zeit in einem Bereich an einem bestimmten Punkt (z.B. beim Leseverstehen von 15-Jährigen wie bei PISA) etwas verändert hat. Auf der Grundlage von wiederholten vergleichbaren Tests und Erhebungen von Kontextdaten über einen längeren Zeitraum können gemessene Veränderungen ins Verhältnis gesetzt und interpretiert werden.

Das Gutachten empfiehlt einen Abstand zwischen Trendmessungen (Erhebungskadenz) von maximal sechs Jahren. Mindestens alle sechs Jahre soll demnach das Erreichen der Grundkompetenzen in derselben Kombination von Fachbereich und Schulstufe überprüft werden. Erst wenn Trends über mehrere Messzeitpunkte hinweg beobachtet werden, können Abweichungen zwischen Messungen vor dem Hintergrund eines Trends interpretiert werden. Nach Einschätzung des Direktors der SKBF sind bereits bei einer Erhebungskadenz von drei Jahren mindestens zwölf Jahre einzuplanen, um eine solide Basis für die Interpretation von Trends bzw. Abweichungen von Trends zu schaffen. Es ist bildungspolitisch bedeutsam, dass neue Trends und bedeutende Abweichungen davon rechtzeitig erfasst werden, damit das

[6] Der PISA-Studie liegt ein Trenddesign zugrunde: Seit dem Jahr 2000 werden alle drei Jahre bei einer repräsentativen Stichprobe von 15-Jährigen die Kompetenzen in den drei Bereichen Mathematik, Schulsprache und Naturwissenschaften gemessen. Indem ein Kernbestand an PISA-Aufgaben immer wieder eingesetzt wird, kann mittels statistischer Methoden zuverlässig erfasst werden, wie sich die Kompetenzen in der Population über die Zeit entwickeln. Bei PISA kommt eine Kontextualisierung mit einem festen Bestand an Fragebogenitems hinzu, anhand derer ebenfalls Trends beobachtet werden können.

Bildungssystem auch rechtzeitig auf unerwünschte Veränderungen reagieren kann. Mit einer hohen Kadenz von Trendmessungen könnten beispielsweise Pandemiefolgen relativ zeitnah erfasst werden.

2.2.2 Kombination von Trenddesign und Kohortenlängsschnitten

Als Variante bzw. Ergänzung zu einem reinen Trenddesign schlägt das Gutachten die Kombination des Trenddesigns mit einem Kohortenlängsschnitt-Design vor. Mit der Kadenz des Trenddesigns (z.B. alle vier Jahre) wird dabei ein Kohortenlängsschnitt gestartet. Dies bedeutet im Fall der ÜGK, dass der gewählte Schülerjahrgang (Kohorte) innerhalb der obligatorischen Schulzeit dreimal an einer ÜGK im selben Fachbereich teilnimmt, nämlich nacheinander im 4., 8. und 11. Schuljahr (Fremdsprachen nur im 8. und 11. Schuljahr). Für jeden der drei Testzyklen werden dabei aus der Kohorte jeweils unabhängige Stichproben gezogen. Von einem Design mit konsequent individuellen Längsschnitten rät die Gutachtengruppe wegen eines ungünstigen Verhältnisses von Aufwand und Nutzen ab. In kleinen Kantonen, in denen alle oder fast alle Schülerinnen und Schüler Teil jeder ÜGK-Stichprobe sind, ergeben sich bei einem Kohortenlängsschnitt-Design von selbst zahlreiche individuelle Längsschnitte und damit Möglichkeiten für Studien zu Aspekten der individuellen Entwicklung. Analysen der Entwicklung von Gruppen innerhalb einer Kohorte sind dank statistischer Methoden auch dann möglich, wenn zu zwei Zeitpunkten aus der Kohorte Zufallsstichproben mit wenig oder keiner Überlappung bei den individuellen Schülerinnen und Schülern gezogen werden.

Gemäss HarmoS-Konkordat (EDK, 2007b, Art. 10) besteht das Ziel des Bildungsmonitorings darin, «systematisch», «kontinuierlich» und «wissenschaftlich gestützt» die «Entwicklungen und Leistungen der obligatorischen Schule» zu evaluieren. Um Entwicklungen der Schule zu verfolgen, eignen sich Trendstudien. Um die Leistungen der Schule zu beurteilen, wird typischerweise analysiert, wie sich Schülerinnen und Schüler bzw. Typen und Gruppen von Schülerinnen und Schülern von einem Messzeitpunkt zum nächsten entwickeln, also z.B. innerhalb eines Zyklus der obligatorischen Schule. Für eine solche Evaluation ist es wichtig, dass jeweils eine Eingangs- und eine Ausgangsmessung vorliegt. Nur so kann die Leistung des betrachteten Teils des Systems tatsächlich beurteilt werden. Wenn die ÜGK also in Form von Kohorten-Längsschnittstudien umgesetzt wird, verfügt sie für das Monitoring über ein entschieden grösseres Potenzial als Trendstudien (wie PISA) für sich allein genommen.

Die Mehrheit des Kosta HarmoS überzeugen die Überlegungen zur Notwendigkeit eines längerfristig festgelegten Trenddesigns und zu den zusätzlichen Vorteilen von Kohorten-Längsschnittstudien, weil diese eine ausgezeichnete Grundlage für fundierte Analysen und Aussagen zur Entwicklung der Qualität und der Harmonisierung des Schulsystems darstellen. Ein Trenddesign mit Kohorten-Längsschnittstudien ist für die Systemevaluation und die Bereitstellung von Erklärungsansätzen besonders geeignet und schafft zudem auch für die Bildungsforschung, die nicht Teil des staatlichen Monitorings ist, eine äusserst interessante Datengrundlage. Dies dürfte letztlich auch der Bildungspolitik und der Bildungssteuerung zugute kommen, denn bekanntlich stützt sich der Bildungsbericht nicht nur auf eigene Analysen, sondern in hohem Masse auch auf Studien, die von anderen Forscherinnen und Forschern erstellt wurden.

Für ein Trenddesign mit Kohorten-Längsschnittstudien spricht auch der Umstand, dass diese Designvariante in der Umsetzung nicht grundsätzlich aufwendiger ist als ein reines Trenddesign. Sie erfordert jedoch mehr Planungsdisziplin.

Unter Abwägung des Kosten-Nutzen-Verhältnisses verschiedener Erhebungsrhythmen erachtet der Kosta HarmoS einen Abstand von vier Jahren zwischen den Trendmessungen als angezeigt. Dieser Rhythmus passt sich auch gut in das übrige nationale Bildungsmonitoring ein: Der Bildungsbericht erscheint alle vier Jahre, ebenso die gemeinsamen bildungspolitischen Ziele (z.B. WBF & EDK, 2019). Zudem wird die PISA-Studie ab 2029 im Vierjahresrhythmus durchgeführt. Bei einer Kadenz von vier Jahren kommen in einem überblickbaren Zeitraum genügend Daten zusammen, um über eine solide Basis für die Interpretation der Trendmessungen zu verfügen. Abstände von vier Jahren sind in der Regel auch nicht zu gross, um wichtige Entwicklungen zu spät festzustellen.

Die folgende Abbildung illustriert ein Kohortenlängsschnitt-Design mit einem Trendabstand (Kadenz) von vier Jahren über einen Zeitraum von 15 Jahren ab der ersten Haupterhebung.

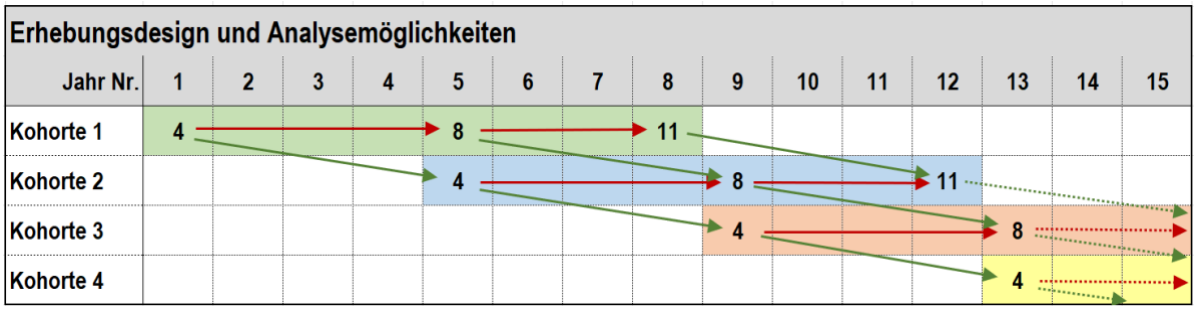


Abb. 1: ÜGK-Erhebungsdesign mit Kohortenlängsschnitten vom 4. bis ins 11. Schuljahr und einer Kadenz von vier Jahren. Analysemöglichkeiten: grüne Pfeile: Trendmessungen über aufeinanderfolgende Kohorten hinweg; rote Pfeile: Veränderungen einer Kohorte innerhalb eines Zyklus der obligatorischen Schule.

Alle vier Jahre beginnt für jede der beiden Fächerkombinationen ein neuer Längsschnitt. Ein Längsschnitt umfasst jeweils drei Erhebungszeitpunkte (4., 8., 11. Schuljahr) mit vorgegebenem Abstand. Bei einer Kadenz von vier Jahren finden die Haupterhebungen im 8. Schuljahr jeweils im selben Jahr statt wie die Haupterhebungen im 4. Schuljahr der darauffolgenden Kohorte. Die Vorarbeiten und die Piloterhebungen können aber flexibel geplant werden. Im dargestellten Erhebungsdesign finden innerhalb der 15 Jahre insgesamt neun Haupterhebungen statt. Dazu kommen unter normalen Umständen ebenfalls neun Piloterhebungen mit sprachregionalen Stichproben sowie rund zweimal so viele Präpilotierungen mit jeweils einigen Klassen pro Sprachregion.

Die durchgezogenen grünen Pfeile in Abb. *1* zeigen, dass innerhalb der 15 Jahre sechs Trendaussagen zwischen aufeinanderfolgenden Zeitpunkten möglich sind. Nicht explizit dargestellt ist, dass auch längere Trends verfolgt werden können, beispielsweise über die vier Messungen im 4. Schuljahr hinweg. Die Zuverlässigkeit von Trendanalysen wächst mit der Länge der Messreihen. Erst wenn einige Messungen vorliegen, können Ausreisser oder Trendwenden als solche identifiziert werden.

Die fünf durchgezogenen roten Pfeile stehen für fünf Möglichkeiten, die Entwicklung der einzelnen Kohorten von einer Schulstufe zur nächsten zu untersuchen. Nicht abgebildet ist die Möglichkeit, die Entwicklungsverläufe verschiedener Kohorten zu vergleichen, z.B. von Kohorte 1 und Kohorte 2, und auch Forschung nach dem Differenz-in-Differenzen-Ansatz zu betreiben, um Ursachen für beobachtete Unterschiede besonders wirkungsvoll zu untersuchen. Das Potenzial eines Datenbestandes für vergleichende Analysen wird laufend grösser, wenn systematisch und dauerhaft Erhebungen durchgeführt werden.

Ein Längsschnittdesign stellt erhöhte Anforderungen an den konzeptionellen Rahmen (Framework) der Tests und der Kontextualisierung, der den Erhebungen zugrunde liegt. Dieser sollte nämlich von Anfang an in Kernbereichen alle drei Erhebungszeitpunkte antizipieren, sodass nicht unnötig Chancen für vertikale Analysen zwischen Klassenstufen verpasst werden. Dies betrifft insbesondere auch die Kompetenzmodelle der Fachbereiche, die möglichst eine klassenstufenübergreifende Entwicklungsdimension aufweisen sollten.

Ein Längsschnittdesign sollte dringend mit der Erweiterung des Spektrums der einzelnen Tests auf den ganzen Kompetenzbereich der Schülerinnen und Schüler (vgl. Abschnitt 2.1.4) einhergehen. Wenn das Kompetenzspektrum unter den Schülerinnen und Schülern zu den einzelnen Zeitpunkten nur ungenügend erfasst werden kann, dann sind auch korrekte vertikale Vergleiche gefährdet. Beispielsweise kann der Werdegang von Gruppen innerhalb einer Kohorte wegen der mangelnden Messgenauigkeit nicht zuverlässig verfolgt werden.

2.2.3 Grad der Standardisierung der Erhebungen

Das Gutachten äussert sich auch zur Frage, ob die Kantone die ÜGK-Erhebungen eigenverantwortlich zu selbst gewählten Zeitpunkten durchführen könnten. Sowohl die befragten Akteurinnen und Akteure als auch die Gutachtengruppe selbst sprechen sich dagegen aus. Das Gutachten weist darauf hin, dass Messungen und Vergleiche wie sie für die ÜGK vorgenommen werden, alterssensibel, zeitfenstersensibel, kohortensensibel und durchführungssensibel sind. Um vergleichbare Daten zu generieren, müssen also zahlreiche Aspekte kontrolliert und in jedem Kanton gleich gehandhabt werden. Zudem könnte bei zeitverschobenen, dezentral administrierten Tests die Testsicherheit nur mit grossem Aufwand gewährleistet werden. Alle Unterlagen müssten sicher verwahrt und sicher zugänglich gemacht werden, wie dies auch bei einem zentralisierten Vorgehen üblich, aber wesentlich einfacher umzusetzen ist.

Für den Kosta HarmoS kommen grundsätzliche Einwände hinzu: Bei der ÜGK handelt es sich nicht etwa um eine Selbstevaluation der einzelnen Kantone, sondern um ein gemeinsames Evaluationsprojekt der beteiligten Kantone, in dessen Rahmen bestimmte Bildungsleistungen der Kantone standardisiert erfasst und verglichen werden. Die genannten Gründe sprechen dafür, von der gegenwärtigen Durchführungspraxis nicht abzuweichen.

Die Projektleitung ÜGK und andere Akteure schlagen anlässlich der Befragungen vor, dass die ÜGK aus Kostengründen in Zukunft unter der Leitung von Lehrpersonen an den schuleigenen Computern durchgeführt wird. Die Gutachtengruppe rät diesbezüglich zur Vorsicht, weil sich in anderen Kontexten Modus-Effekte gezeigt haben: Der Wechsel von externen Testleitenden zu Lehrpersonen hat die Ergebnisse bei bestimmten Aufgabentypen beeinflusst. Insbesondere wird davon abgeraten, Piloterhebung und Haupterhebung unter unterschiedlichen Bedingungen durchzuführen.

2.3 Weitere Aspekte der Umsetzung der ÜGK

Das Gutachten äussert sich zu einer ganzen Reihe von fachlich-technischen Aspekten der ÜGK und gibt entsprechende Einschätzungen und Empfehlungen ab. Es handelt sich um die folgenden Punkte:

- Testentwicklung und Schwellenwertsetzung
  - Fachdidaktische Verankerung
  - Konzeptionelle und konkrete Grundlagen für die Testentwicklung
  - Passung der Aufgabenschwierigkeit
  - Aufgabenübersetzungen
  - Validierungsstudien
  - Timing des Testentwicklungsprozesses
  - Schaffung von Aufgabenpools
  - Schwellenwertsetzung
- Planung und Umsetzung der Kontextualisierung (vor allem Fragebögen)
- Datennutzung für Sekundäranalysen
- Dokumentation der ÜGK-Zyklen
- Berichterstattung und Kommunikation.

Der Kosta HarmoS hat sich mit diesen Punkten auseinandergesetzt. Weil es sich weitgehend um eine fachlich-technische Diskussion handelt, die sich primär an die Umsetzungsebene richtet, wird sie an dieser Stelle nicht wiedergegeben. Sie kann aber in Anhang 1 nachgelesen werden.

2.4 Governance der ÜGK

Wie in Abschnitt 1.2.3 erläutert, hat die ÜGK erst im Hinblick auf die ÜGK-Zyklen 2020 und 2022 (bzw. 2023 und 2024) die in den Rechtsgrundlagen vorgesehene Struktur erhalten. Darin nimmt die Universität

Bern bzw. dessen Interfakultäres Zentrum für Bildungsforschung ICER auf der operationellen Ebene wesentliche Koordinations- und Leitungsfunktionen wahr. Das Gutachten würdigt die Neugestaltung der Strukturen als Fortschritt, identifiziert aber weiteren Verbesserungsbedarf im Bereich der Governance.

2.4.1 Abgrenzung zwischen Bildungspolitik bzw. Bildungsverwaltung und Wissenschaft

Die Gutachtengruppe erinnert daran, dass die Wissenschaft in Bezug auf die ÜGK zwei unterschiedliche Rollen einnimmt: 1) wissenschaftliche Dienstleistung, 2) Beteiligung an der ÜGK aus eigenem Interesse (insbesondere an der Kontextualisierung oder durch Sekundäranalysen von ÜGK-Daten). Dabei ist das Interesse der Wissenschaft im Fall der ÜGK den Interessen der Bildungspolitik grundsätzlich nachgelagert. Die Umsetzung wissenschaftlicher Arbeiten aber soll ohne politische Einflussnahme allein nach wissenschaftlichen Standards erfolgen.

Die Gutachtengruppe sieht die folgenden Aufgaben, die aktuell Bildungspolitik und Bildungsverwaltung wahrnehmen, in der Zuständigkeit der Wissenschaft:

- Festlegung der Schwellenwerte: Gemäss ÜGK-Reglement beschliesst die EDK-Plenarversammlung die Schwellenwerte und hat damit die Möglichkeit, auf die Höhe des Anteils der Schülerinnen und Schüler, welche die Grundkompetenzen erreichen, Einfluss zu nehmen. Die technische Schwellenwertsetzung, aus der die Vorschläge für die Schwellenwerte hervorgehen, liegt zudem in der Zuständigkeit der ADB, die wiederum Teil der EDK ist. Für die Gutachtengruppe ist die Schwellenwertsetzung eine wissenschaftliche Aufgabe und sollte nicht durch eine Verwaltungsstelle durchgeführt werden.
- Genehmigung des Kontextfragebogens: Der ÜGK-Kontextfragebogen für die Schülerinnen und Schüler wird gemäss aktuellen Rechtsgrundlagen dem Kosta HarmoS zur Konsultation vorgelegt und vom EDK-Vorstand genehmigt. Diese Praxis hat sich etabliert, weil vor allem die Fragen zum sozialen Hintergrund als heikel und den aktuellen Verhältnissen nicht immer angemessen empfunden wurden. Das Gutachten erachtet die Fragebogenentwicklung als rein wissenschaftliche Aufgabe und schlägt vor, dass anstelle einer abschliessenden Genehmigung des Fragebogens zu Beginn der Arbeiten an der Kontextualisierung zwischen den Auftraggebenden und den Auftragnehmenden ein Rahmenkonzept (Framework) ausgehandelt und verbindlich beschlossen wird.
- Anbindung der ADB an das GS EDK: Die ADB ist zwar für die (wissenschaftlichen) Kernbereiche der Aufgaben- und Testentwicklung sowie der Schwellenwertsetzung zuständig, aber wissenschaftlich kaum in die ÜGK eingebunden. Diese Positionierung schwächt gemäss Gutachten nicht nur die Glaubwürdigkeit der Schwellenwertsetzung, sondern ermöglicht dem wissenschaftlichen Personal wenig wissenschaftlichen Austausch und gibt ihm wenig Perspektiven. Zudem birgt die knappe personelle Dotierung (früher zwei Personen, heute eine) beträchtliche Risiken, was den Knowhow-Verlust angeht. Das Gutachten empfiehlt eine Loslösung der ADB vom Generalsekretariat der EDK und eine Eingliederung ihrer Funktionen in die wissenschaftlichen Strukturen.

Der Kosta HarmoS unterstützt eine klare funktionale Trennung zwischen der Ebene der Bildungspolitik bzw. Bildungsadministration und der Ebene der wissenschaftlichen Umsetzung der ÜGK. Die Einigung auf ein Framework, das ausgehend von den Zielen der ÜGK bzw. einer ÜGK-Erhebung die Eckwerte vor allem in den Bereichen Stichprobenziehung, Testentwicklung und Kontextualisierung verbindlich regelt, scheint dafür ein geeignetes Instrument zu sein. Nach der gegenseitigen Verpflichtung auf ein Framework führt die auftragnehmende wissenschaftliche Institution die Entwicklungsarbeiten zusammen mit ihren Vertragspartnern nach wissenschaftlichen Standards aus und ist mit dem Ergebnis gegenüber der EDK rechenschaftspflichtig.

Der Kosta HarmoS erachtet auch die Integration der Funktionen der ADB, die mit der ÜGK zu tun haben, in die wissenschaftlichen Strukturen der ÜGK für sinnvoll.

Gleichzeitig schlägt er vor, auf die Funktionen, welche die ADB gemäss Reglement zwar hat, aber heute nicht wahrnehmen muss, zu verzichten. Dabei handelt es sich um Folgendes:

a) Bereitstellung einer multifunktionalen und mehrsprachigen technischen Struktur,

b) Sicherstellung einer gesamtschweizerischen Messskala (nicht nur für die ÜGK),
c) Koordination des Austauschs von Aufgaben für die Überprüfung der Erreichung der nationalen Bildungsziele (Grundkompetenzen) sowie für Projekte der Kantone.

Der Kosta HarmoS weist darauf hin, dass das Bedürfnis nach einer gemeinsamen Messskala und einem gesamtschweizerischen Austausch von Testaufgaben grundsätzlich weiterbesteht, auch wenn entsprechende Leistungen der ADB in den letzten Jahren nicht nachgefragt wurden..

2.4.2 «Diffusion» in der Leitungsfunktion des ICER und Schnittstellenrisiken

Die Kooperationsvereinbarung zwischen der EDK und der Universität Bern betreffend die Durchführung der beiden geplanten ÜGK-Zyklen gibt dem ICER im ÜGK-Projekt teils eine Koordinations- und teils eine Steuerungs- und Leitungsfunktion. Gleichzeitig trägt die Universität Bern die Verantwortung für das Ergebnis. Das Gutachten nennt dies eine «Diffusion» der Leitungsfunktion. Zwei Aspekte sind dabei zu unterscheiden. Der erste Aspekt betrifft die vertraglichen Beziehungen zwischen den Institutionen, die für die Umsetzung der ÜGK zuständig sind: Die Universität Bern schliesst zwar mit den meisten der beteiligten Institutionen IBE, PHSG, SRED, SUPSI und für die ÜGK 2024 auch mit der FHGR selbständig Kooperationsvereinbarungen ab, ist aber für einen erfolgreichen Projektabschluss besonders auch auf die inhaltlich zentralen Beiträge der ADB angewiesen, die zum GS EDK gehört. Aus Sicht der Gutachtengruppe ist dadurch vor allem die Universität Bern Schnittstellenrisiken ausgesetzt, die elminiert werden sollten. Sie empfiehlt deshalb auch aus diesem Grund eine Einordnung der Funktionen der ADB (und der FHGR) in die Projektstruktur unter der Leitung des ICER. Der zweite Aspekt betrifft die Leitungsfunktion des ICER in der PL ÜGK, die Aspekte der Kooperation, Koordination, Steuerung und Leitung aufweist. Das Gutachten empfiehlt, in der Projektleitung der ÜGK generell klare, vereinfachte Zuständigkeits-, Entscheidungs- und Verantwortungsstrukturen zu implementieren, die auch Konflikten oder Zeitdruck standhalten können. Das Gutachten weist darauf hin, dass der Aufbau einer effizienten Organisationsstruktur auch davon abhängt, dass eine langfristige Planung die Eckwerte vorgibt, beispielsweise den Erhebungsrhythmus. Das Ziel sollte aus Sicht der Gutachtengruppe ein stabiles Kompetenzzentrum mit stabilem Knowhow sein.

Der Kosta HarmoS teilt die Einschätzung, dass unklare oder nur partielle Zuständigkeiten Schnittstellenrisiken bergen und begrüsst auch deshalb die Integration der Funktionen der ADB in die wissenschaftlichen Strukturen. Eine stärkere Hierarchisierung der Beziehungen zwischen dem ICER und den übrigen Institutionen, die heute in der PL ÜGK vertreten sind, muss sorgfältig geprüft werden. Dabei ist u.a. Folgendes zu beachten:

- Es ist der EDK ein Anliegen, dass die ÜGK innerhalb des eigenen Bildungssystems durchgeführt wird, u.a. weil die entsprechenden Kompetenzen aufgebaut und erhalten werden sollen.
- Es besteht ein politischer Wille, dass die drei Sprachregionen bei der ÜGK zusammenarbeiten. Ziffer 3.2 der Vereinbarung zwischen der EDK und der Universität Bern verpflichtet zum Beispiel das ICER, mit den Sprachregionen zusammenzuarbeiten.
- Die Zusammenarbeit von Institutionen aus allen Sprachregionen erscheint nicht nur wünschbar, sondern auch notwendig. Die ÜGK mit ihren grossen Stichproben muss ein grosses Praxisfeld nachhaltig betreuen und motivieren. Ohne regionale Verwurzelung ist das schwer vorstellbar.
- Der Furcht vor Schnittstellenrisiken ist entgegenzuhalten, dass die PL ÜGK bisher offenbar Formen gefunden hat, um erfolgreich zusammenzuarbeiten.

Der Kosta HarmoS hat im Zusammenhang mit der Governance die Umsetzung des wissenschaftlichen Konsortiums Bildungsforschung und Evaluation (WIKO BFE, vgl. Abschnitt 1.2.3), die seit den Anfangszeiten der ÜGK ein Thema ist, und die gleichwertige Berücksichtigung der Sprachregionen in der ÜGK diskutiert. Er wird dazu Vorschläge unterbreiten.

2.4.3 Stellung der Fachdidaktiken

Die Fachdidaktiken sind bisher nur indirekt in die Projektstrukturen integriert, d.h. sie sind nicht Teil der thematischen Arbeitsgruppen der PL ÜGK. In der Zusammenarbeit mit der ADB leisten sie einen wesentlichen Teil der Aufgabenentwicklung, beteiligen sich an den Workshops zur Schwellenwertsetzung und leisten Beiträge zur Berichterstattung. Wie andere Forschende haben sie zudem die Möglichkeit, eigene Interessen (Konstrukte) in die Kontextualisierung einzubringen. Die Gutachtengruppe empfiehlt, die Fachdidaktiken umfassender als bisher in das Projekt zu integrieren und sie gleichzeitig im Hinblick auf ihre Aufgaben in der ÜGK zu professionalisieren. Durch die Einbindung der Fachdidaktiken soll u.a. die Daten- und Ergebnisnutzung auch eine fachdidaktische Komponente erhalten, was sich günstig auf die Aus- und Weiterbildung der Lehrpersonen und den Unterricht auswirken und die Wirksamkeit der Bildungsziele unterstützen kann.

Aus Sicht des Kosta HarmoS ist eine umfassendere und zentralere Einbindung der Fachdidaktik in die wissenschaftlichen Strukturen der ÜGK ein längst fälliger Schritt. Zudem ist sie eine konsequente Ergänzung der Neuzuordnung der ÜGK-bezogenen Funktionen der ADB. Die Aufgaben- und Testentwicklung soll weiterhin der wichtigste Tätigkeitsbereich der Fachdidaktiken sein. Da sollen sie weiter gestärkt werden (vgl. dazu Anhang 1, Abschnitt «Testentwicklung, Schwellenwertsetzung»). Zudem soll ihnen in Zukunft eine wichtige Funktion bei der Erarbeitung und periodischen Erneuerung des konzeptionellen Rahmens (Framework) der jeweiligen ÜGK-Erhebungen zukommen.

3 Vorgehen im Hinblick auf die Weiterentwicklung der ÜGK

Der Kosta HarmoS begleitet die Arbeiten der ÜGK seit den Anfängen. Er wurde von den Gremien der EDK eingesetzt, um die notwendigen Grundlagen zu schaffen, die Umsetzung zu begleiten und entsprechende Entscheidgrundlagen zu erarbeiten. Er hat sich im Auftrag der Gremien der EDK sowie auf der Grundlage der vorliegenden Gutachten und im Dialog mit weiteren Experten intensiv mit der Beurteilung der bisherigen Umsetzung der ÜGK beschäftigt. Im Hinblick auf die Weiterführung der ÜGK hat er Befunde und Empfehlungen zur Kenntnis genommen und verschiedene Optionen geprüft.

Die grosse Mehrheit der Mitglieder des Kosta HarmoS befürwortet eine Weiterführung und Weiterentwicklung der ÜGK. Für einzelne Mitglieder kommt ein gänzlicher Verzicht auf die ÜGK und stattdessen eine Ausweitung der PISA-Studie auf kantonal repräsentative Stichproben der Schülerinnen und Schüler des 11. Schuljahres in Frage, obschon das HarmoS-Konkordat von 2007 die Überprüfung des Erreichens der Grundkompetenzen als Steuerungs- und Harmonisierungsinstrument explizit vorsieht.

Im Hinblick auf die Beschlussfassung zur Weiterführung der ÜGK schlägt der Kosta HarmoS den Gremien der EDK in einem ersten Schritt Eckwerte für ein «Monitoring der Grundkompetenzen» vor. Die Eckwerte umfassen das zukünftige Erhebungsdesign und verschiedene Governance-Aspekte (Institutionen, Strukturen, Zuständigkeiten). Die vorgeschlagenen Neuerungen beinhalten verschiedene Verbesserungen an der aktuellen Praxis der ÜGK. Das Ziel der vorgelegten Eckwerte zur Weiterführung der ÜGK ist die Steigerung des Werts und des Nutzens der ÜGK für das Bildungsmonitoring und auf diesem Weg für die evidenzbasierte Steuerung der Bildungssysteme in der Schweiz.

Sobald die Eckwerte beschlossen sind, sollen die Rechtsgrundlagen für die Umsetzung des «Monitorings der Grundkompetenzen» angepasst und den Gremien der EDK vorgelegt werden.

3.1 Neues Erhebungsdesign

Die bisherigen und bis 2024 geplanten vier Haupterhebungen im Rahmen der ÜGK haben den Charakter von erstmaligen Einzelmessungen. Ein längerfristig festgelegtes Erhebungsdesign mit einem gezielt festgelegten Erhebungsrhythmus und geplanten Bezügen zwischen den einzelnen Erhebungen bietet im Vergleich zu solchen Querschnittstudien wesentliche Vorteile: Es erlaubt, Entwicklungstrends aufzuzeigen und die Leistungen der obligatorischen Schule aussagekräftiger zu evaluieren. Zudem wird der Beitrag der ÜGK zum nationalen und kantonalen Bildungsmonitoring berechenbar und damit für eine Reihe von Akteuren relevanter.

Der Kosta HarmoS hat sich in vier Sitzungen vom 24. Januar bis zum 28. September 2022 mit verschiedenen Varianten des Erhebungsdesigns auseinandergesetzt und legt dem Vorstand vier Varianten vor. Alle Varianten sehen ein Erhebungsdesign vor, das auf langfristiger Planung und Regelmässigkeit basiert.

Die Design-Varianten unterscheiden sich in Bezug auf vier Merkmale:

- Getestete **Fachbereiche**: Bei der ÜGK kommen im Maximum die vier Fachbereiche mit freigegeben Grundkompetenzen in Frage, d.h. Schulsprache, Mathematik, Fremdsprachen, Naturwissenschaften. Mit Blick auf die Kosten sehen alle Varianten den Verzicht auf die Überprüfung der Grundkompetenzen in den Naturwissenschaften vor.
  Bei PISA werden obligatorisch die drei Fachbereiche Schulsprache (Lesen), Mathematik und Naturwissenschaften (gemäss Definition der OECD) getestet, dies bei 15-Jährigen.
  Zwei Design-Varianten schlagen vor, dass alle Kantone mit jeweils einer repräsentativen Stichprobe des 11. Schuljahres an PISA teilnehmen sollen. Seit 2015 beteiligt sich die Schweiz mit einer nationalen Stichprobe der 15-Jährigen an PISA. Dies wird unabhängig von den Varianten (auch) beibehalten.

- Die **Klassenstufen**, auf denen getestet werden soll (4., 8. und/oder 11. Schuljahr).
- **Kadenz** der Erhebungen: Ausgangspunkt bei allen Varianten ist der Vierjahreszyklus der Bildungsberichterstattung und der PISA-Studien ab 2029. Variante B sieht bei den Fremdsprachen eine 8-jährige Kadenz vor.
- **Kohortenstudien**: Bei allen Varianten A-D können wegen der Wiederholung der einzelnen Messungen im Vierjahresrhythmus Trends beobachtet werden. Jedoch unterscheiden sich die Varianten darin, ob und wie lange die Entwicklung einzelner Schülerjahrgänge in der obligatorischen Schulzeit verfolgt werden kann (Kohorten-Längsschnittstudien). Bei Variante D kann die Entwicklung von Kohorten nicht verfolgt werden. Damit entfällt die Möglichkeit, die Leistungen der einzelnen Schulzyklen zu evaluieren.

Tabelle 1: Varianten des Erhebungsdesigns

	Aktuelle ÜGK	**Variante A**	**Variante B**	**Variante C**	**Variante D**
Fachbereiche HarmoS	Schulsprache Mathematik Fremdsprachen	Schulsprache Mathematik Fremdsprachen	Schulsprache (4/8) Mathematik (4/8) Fremdsprachen (8/11)	Schulsprache Mathematik Fremdsprachen	-
Fachbereiche PISA (Zusatzstichprobe 11. Schuljahr)	-	-	Schulsprache Mathematik Naturwissenschaften	-	Schulsprache Mathematik Naturwissenschaften
Stufen	4/8/11	4/8/11	4/8/11	8/11	11
Kadenz	offen	4 Jahre	4 Jahre Fremdsprachen 8 J.	4 Jahre	4 Jahre
Kohortenstudien	-	4→8→11	4→8→11	8→11	-
Anzahl ÜGK-Erhebungen in 15 Jahren	offen	9	10	7	-
Anzahl PISA-Erhebungen im 11. Schuljahr in 15 Jahren	-	-	4	-	4
Kosten ÜGK pro Jahr Ø 2025-2041*	CHF 1.5 Mio.	CHF 2.25 Mio.	CHF 2.3 Mio.	CHF 1.95 Mio.	-
Mehrkosten PISA 11. Schuljahr pro Jahr	-	-	CHF 0.5 Mio.	-	CHF 0.5 Mio.
Kosten PISA 15-Jährige pro Jahr	CHF 0.6 Mio.	CHF 0.6 Mio.	CHF 0.6 Mio.	CHF 0.6 Mio.	CHF 0.6 Mio.
Kosten Total pro Jahr	CHF 2.1 Mio.	CHF 2.85 Mio.	CHF 3.4 Mio.	CHF 2.55 Mio.	CHF 1.1 Mio.

* Als Planungsgrundlage dienen die Haupterhebungen im 15-jährigen Zeitraum 2027-2041. Soll 2027 eine Haupterhebung durchgeführt werden, müssen die Vorarbeiten spätestens im Jahr 2025 beginnen.

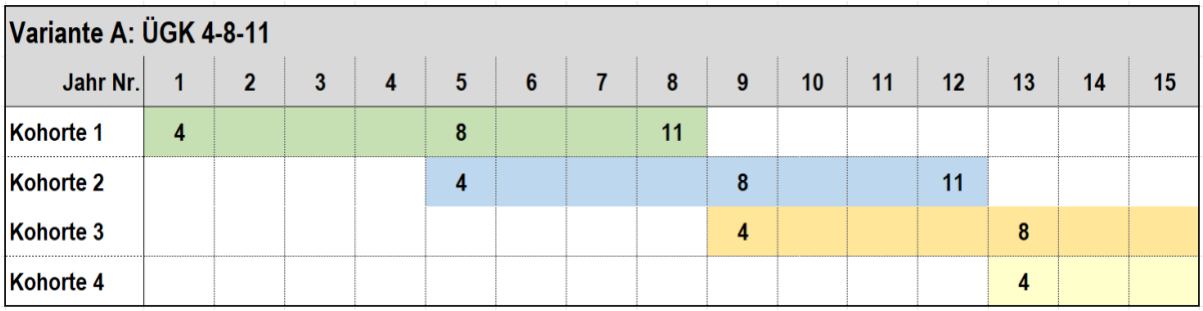
3.1.1 Die Varianten

Im Folgenden werden die Varianten A bis D kurz charakterisiert und anschliessend anhand von besonders relevanten Kriterien beurteilt, soweit dies möglich ist. Innerhalb des Kosta HarmoS werden die Belastungen durch die verschiedenen Erhebungsdesigns unterschiedlich eingeschätzt, weshalb bezüglich Belastungen auf eine Beurteilung verzichtet wird.

Die Varianten A und C geniessen im Kosta HarmoS deutlich mehr Unterstützung als die Varianten B und D. Insgesamt wird Variante C zur Umsetzung empfohlen.

- **Variante A**

Variante A erhebt gegen Ende des 4., 8. und 11. Schuljahres im Rahmen von ÜGK-Erhebungen Leistungs- und Kontextdaten. Auf den drei Klassenstufen wird in den Fächern Schulsprache, Mathematik und Fremdsprachen insbesondere das Erreichen der Grundkompetenzen überprüft. Die Naturwissenschaften werden nicht berücksichtigt.



Die Messungen zu den drei Zeitpunkten in derselben Kohorte erlauben es, im Sinne des Bildungsmonitorings die Leistungen des zweiten und dritten Zyklus der obligatorischen Schule zu evaluieren, weil für diese beiden Zyklen Eingangs- und Ausgangsmessungen vorgenommen werden.

Alle vier Jahre (Kadenz) beginnen die Erhebungen im 4. Schuljahr mit einer neuen Kohorte. Damit ist es möglich, für das 4., 8. und 11. Schuljahr über die Jahre hinweg Entwicklungstrends zu erkennen und zu verfolgen.

Unabhängig davon beteiligt sich die Schweiz, wie seit 2015 üblich, lediglich mit einer national repräsentativen Stichprobe der 15-Jährigen an der PISA-Studie. Dort werden die «Kerndomänen» Schulsprache (Lesen), Mathematik und Naturwissenschaften getestet.

Globale Einschätzung: Die Variante A setzt im Vergleich zu den anderen Varianten die mit dem HarmoS-Konkordat gesetzten Ziele am besten um. Die Aussagekraft der Resultate ist aufgrund des Testens aller Fachbereiche ausser Naturwissenschaften auf allen in Frage kommenden Stufen hoch.

Tabelle 2: Beurteilung der Variante A nach einzelnen Kriterien

Kriterium	**Beurteilung**
HarmoS-Konkordat 2007: Überprüfung des Erreichens der Bildungsziele, die der Harmonisierung dienen	+ Überprüfung der Bildungsziele Schulsprache + Überprüfung der Bildungsziele Mathematik + Überprüfung der Bildungsziele Fremdsprachen – keine Überprüfung Bildungsziele Naturwissenschaften
HarmoS-Konkordat 2007: Beitrag zu einer systematischen und kontinuierlichen Evaluation der Entwicklungen und Leistungen des Systems	+ Für alle dafür in Frage kommenden Zyklen der obligatorischen Schule Eingangs- und Ausgangsmessungen umgesetzt: • Trendmessungen zu Entwicklungen im 4., 8. und 11. Schuljahr • Evaluation der Leistungen des 2. und 3. Zyklus

Kriterium	Beurteilung
Nutzen für - Bildungsmonitoring - Bildungssteuerung - Bildungsforschung	+ Schaffung einer umfassenden Datengrundlage zum 2. und 3. Zyklus der obligatorischen Schule (Kompetenzen und Kontext) • umfassende Eingangs- und Ausgangsmessungen für Trend- und Kohortenstudien • regelmässige Datenlieferungen • Möglichkeit, eigene Interessen in Kontextualisierung einzubringen • mit anderen Datenbeständen der öffentlichen Hand verknüpfbar
Aufbau und Erhalt von Kompetenzen im Bereich Large-Scale-Assessments	+ Permanente professionelle Strukturen sind Bedingung für erfolgreiche Umsetzung – Fachdidaktik der Naturwissenschaften ausgeschlossen
Belastung der Schulen und Verwaltungen im Zusammenhang mit den Erhebungen	Innerhalb von 15 Jahren gibt es: • 9 ÜGK-Haupterhebungen, 9 Piloterhebungen, max. 18 Prä-pilotierungen • 4 PISA-Haupterhebungen (bei nationalen Stichproben), dazu 4 Pilot-erhebungen (*field trials*)
Durchschnittliche Kosten pro Jahr über 15 Jahre ab der ersten Haupterhebung (bei der ÜGK zusätzlich 2 Jahre Vorbereitung)	• ÜGK ca. CHF 2.25 Mio. pro Jahr (Projektkosten) • PISA keine Zusatzkosten, da nur nationale Stichprobe der 15-J. (Kosten dafür 50% vom Bund getragen; Anteil EDK ca. 0.6 Mio. pro Jahr)

- **Variante B**

Bezeichnend für Variante B ist das Ineinandergreifen von ÜGK-Erhebungen und PISA-Erhebungen mit kantonal repräsentativen Stichproben von Schülerinnen und Schülern des 11. Schuljahres (wie bei der ÜGK im 11. Schuljahr).

Variante B: ÜGK und PISA 11. Schuljahr verzahnt

Jahr Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kohorte 1			PISA												
Kohorte 2				8			PISA								
Kohorte 3				4				8			PISA				
Kohorte 4								4				8			PISA
Kohorte 5												4			
Kohorte 1 FS		8			11										
Kohorte 2 FS										8			11		

Im 4. und 8. Schuljahr finden mit einer Kadenz von vier Jahren ÜGK-Erhebungen in den Fachbereichen Schulsprache und Mathematik statt. Im 11. Schuljahr nimmt jeweils dieselbe Kohorte mit kantonal repräsentativen Stichproben an der PISA-Studie teil. Mit einer Kadenz von 8 Jahren werden unabhängig davon, in anderen Kohorten, im 8. und 11. Schuljahr die Fremdsprachenkompetenzen gemessen.

Mit diesem Design können wie in Variante A die Leistungen des zweiten und dritten Zyklus im Rahmen des Bildungsmonitorings evaluiert werden. Beim Überprüfen der Grundkompetenzen wird aber nicht nur auf die Naturwissenschaften verzichtet, sondern am Ende der obligatorischen Schulzeit auch auf die Fachbereiche Schulsprache und Mathematik. Der Einbezug der PISA-Studie in die Messreihe hat zudem Konsequenzen

für die Kontextualisierung: Die Kontextfragebogen müssen auf PISA abgestimmt werden, um ihren Zweck zu erfüllen (v.a. Verfolgen von Veränderungen). Wie die Kontextualisierung im 8. und 11. Schuljahr bei den Fremdsprachen aussehen wird, ist offen.

Die Variante B erlaubt, ähnlich wie Variante A, auch Trendaussagen für das 4., 8. und 11. Schuljahr. Bei den Fremdsprachen liegen allerdings die Messungen mit einem Abstand von acht Jahren für diesen Zweck sehr weit auseinander.

An der PISA-Studie beteiligt sich die Schweiz bei Variante B einerseits neu mit kantonal repräsentativen Stichproben der Schülerinnen und Schüler des 11. Schuljahres und andererseits wie in den vergangen Jahren mit einer national repräsentativen Stichprobe der 15-Jährigen.

Globale Einschätzung: Die Variante B führt zu Erkenntnissen in Bezug auf vier Fachbereiche. Die Anforderungen entsprechen aber zum Teil den Vorstellungen der OECD. Am Ende der obligatorischen Schulzeit sind ausser bei den Fremdsprachen keine Erkenntnisse zum Erreichen der definierten Bildungsziele möglich. Variante B ist aufgrund der Verzahnung von ÜGK und PISA kompliziert und anspruchsvoll in Umsetzung und Kommunikation.

Tabelle 3: Beurteilung der Variante B nach einzelnen Kriterien

Kriterium	**Beurteilung**
HarmoS-Konkordat 2007: Überprüfung des Erreichens der Bildungsziele, die der Harmonisierung dienen.	± Überprüfung der Bildungsziele Schulsprache im 4. und 8. Schuljahr; keine Überprüfung im 11. Schuljahr, stattdessen PISA ± Überprüfung der Bildungsziele Mathematik im 4. und 8. Schuljahr; keine Überprüfung im 11. Schuljahr, stattdessen PISA – Keine Überprüfung der Bildungsziele Naturwissenschaften; im 11. Schuljahr PISA + Überprüfung der Bildungsziele Fremdsprachen, wenn auch mit tiefer Kadenz (8 Jahre)
HarmoS-Konkordat 2007: Beitrag zu einer systematischen und kontinuierlichen Evaluation der Entwicklungen und Leistungen des Systems	+ Für alle dafür in Frage kommenden Zyklen der obligatorischen Schule Eingangs- und Ausgangsmessungen umgesetzt: • Trendmessungen zu Entwicklungen im 4., 8. und 11. Schuljahr • Evaluation der Leistungen des 2. und 3. Zyklus
Nutzen für - Bildungsmonitoring - Bildungssteuerung - Bildungsforschung	+ Schaffung einer umfassenden Datengrundlage zum 2. und 3. Zyklus der obligatorischen Schule (Kompetenzen und Kontext) • umfassende Eingangs- und Ausgangsmessungen für Trend- und Kohortenstudien • regelmässige Datenlieferungen • Orientierung an der Kontextualisierung der PISA-Studie • mit anderen Datenbeständen der öffentlichen Hand verknüpfbar – Wegen der Abstimmung mit PISA im 11. Schuljahr ist es kaum möglich, eigene Items und Interessen in Test und/oder Kontextualisierung einzubringen. ± Die Überprüfung der Bildungsziele Fremdsprachen hat zwar eine tiefe Kadenz, könnte aber die Möglichkeit bieten, eigene Interessen in die Kontextualisierung einzubringen
Aufbau und Erhalt von Kompetenzen und Strukturen im Bereich Large-Scale-Assessments	+ Permanente professionelle Strukturen sind Bedingung für erfolgreiche Umsetzung – Fachdidaktik der Naturwissenschaften ausgeschlossen

Kriterium	Beurteilung
Belastung der Schulen und Verwaltungen im Zusammenhang mit den Erhebungen	Innerhalb von 15 Jahren gibt es: • 10 ÜGK-Haupterhebungen, 10 Piloterhebungen, max. 20 Prä-pilotierungen • 4 PISA-Haupterhebungen mit Stichproben, die umfangreicher sind als ÜGK-Stichproben im 11. Schuljahr; dazu 4 Piloterhebungen (*field trials*)
Durchschnittliche Kosten pro Jahr über 15 Jahre ab der ersten Haupterhebung (bei der ÜGK zusätzlich 2 Jahre Vorbereitung)	• ÜGK ca. CHF 2.3 Mio. pro Jahr (Projektkosten) • PISA zusätzlich ca. CHF 0.5 Mio. pro Jahr für kantonal repräsentative Stichproben; (Kosten für nationale Stichprobe 50% vom Bund getragen; Anteil EDK ca. 0.6 Mio. pro Jahr)

- **Variante C**

Das Design der Variante C entspricht weitgehend der Variante A, sieht aber im 4. Schuljahr keine Erhebungen vor. Diese beschränken sich auf das 8. und 11. Schuljahr. Dort wird in den Fachbereichen Schulsprache, Mathematik und Fremdsprachen insbesondere das Erreichen der Grundkompetenzen überprüft. Die Naturwissenschaften werden nicht berücksichtigt.

Variante C: ÜGK 8-11

Jahr Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kohorte 1	8			11											
Kohorte 2					8			11							
Kohorte 3									8			11			
Kohorte 4													8		

Auf Grundlage der Messungen im 8. und 11. Schuljahr derselben Kohorte können lediglich die Leistungen des dritten Zyklus (Sek I) der obligatorischen Schule im Sinne des Bildungsmonitorings evaluiert werden. Die Ausgangsmessung für den ersten Zyklus, die gleichzeitig auch Eingangsmessung für den zweiten Zyklus ist, fehlt.

Auch bei Variante C beginnen die Erhebungen alle vier Jahre mit einer neuen Kohorte, dies allerdings erst im 8.Schuljahr. Damit können für das 8. und 11. Schuljahr über die Jahre hinweg Entwicklungstrends verfolgt werden.

Unabhängig von der ÜGK beteiligt sich die Schweiz, wie in Variante A, weiterhin mit einer national repräsentativen Stichprobe der 15-Jährigen an der PISA-Studie.

Globale Einschätzung: Die Variante C setzt die mit dem HarmoS-Konkordat gesetzten Ziele im Vergleich zur Variante A mit klaren Abstrichen um. Die Aussagekraft für das 8. und 11. Schuljahr sowie den dritten Zyklus (Sek I) ist dank dem Testen von drei Fachbereichen hoch. Der Verzicht auf die ÜGK im 4. Schuljahr führt aber zu einer empfindlichen Schmälerung der Aussagekraft dadurch, dass die Ausgangsmessung für den ersten und damit die Eingangsmessung für den zweiten Zyklus fehlt.

Tabelle 4: Beurteilung der Variante C nach einzelnen Kriterien

Kriterium	**Beurteilung**
HarmoS-Konkordat 2007: Überprüfung des Erreichens der Bildungsziele, die der Harmonisierung dienen.	± Überprüfung der Bildungsziele Schulsprache im 8. und 11. Schuljahr ± Überprüfung der Bildungsziele Mathematik im 8. und 11. Schuljahr ± Überprüfung der Bildungsziele Fremdsprachen im 8. und 11. Schuljahr – Das Erreichen der Bildungsziele im 4. Schuljahr wird nicht überprüft.
HarmoS-Konkordat 2007: Beitrag zu einer systematischen und kontinuierlichen Evaluation der Entwicklungen und Leistungen des Systems	± Nur ein Teil der Trendaussagen sowie Aussagen zu den Leistungen der Zyklen möglich: • Trendmessungen zu Entwicklungen im 8. und 11. Schuljahr • Evaluation der Leistungen des 3. Zyklus
Nutzen für - Bildungsmonitoring - Bildungssteuerung - Bildungsforschung	± Schaffung einer umfassenden Datengrundlage insbesondere zum 3. Zyklus der obligatorischen Schule (Kompetenzen und Kontext): • Eingangs- und Ausgangsmessungen für Trend- und Kohortenstudien in zwei Schuljahren • regelmässige Datenlieferungen • Möglichkeit, eigene Interessen in Kontextualisierung einzubringen • mit anderen Datenbeständen der öffentlichen Hand verknüpfbar
Aufbau und Erhalt von Kompetenzen im Bereich Large-Scale-Assessments Aufbau und Erhalt von Kompetenzen im Bereich Large-Scale-Assessments	+ Permanente professionelle Strukturen sind Bedingung für erfolgreiche Umsetzung – Fachdidaktik der Naturwissenschaften ausgeschlossen
Belastung der Schulen (und der Umsetzungsebene der kt. Verwaltungen) im Zusammenhang mit den Erhebungen	Innerhalb von 15 Jahren gibt es: • 7 ÜGK-Haupterhebungen, 7 Piloterhebungen, max. 14 Präpilotierungen • 4 PISA-Haupterhebungen (nationale Stichproben), dazu 4 Pilot-erhebungen (*field trials*)
Durchschnittliche Kosten pro Jahr über 15 Jahre ab der ersten Haupterhebung (bei der ÜGK zusätzlich 2 Jahre Vorbereitung)	• ÜGK ca. CHF 1.95 Mio. pro Jahr • PISA keine Zusatzkosten, da nur nationale Stichprobe der 15-J. (Kosten dafür 50% vom Bund getragen; Anteil EDK ca. 0.6 Mio. pro Jahr)

- **Variante D**

Variante D verzichtet ganz auf die ÜGK. Die Schweiz beteiligt sich lediglich an der PISA-Studie, und zwar einerseits mit kantonal repräsentativen Stichproben der Schülerinnen und Schüler des 11. Schuljahres und andererseits weiterhin mit einer national repräsentativen Stichprobe der 15-Jährigen.

Variante D: PISA 11. Schuljahr															
Jahr Nr.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Kohorte 1			PISA												
Kohorte 2							PISA								
Kohorte 3											PISA				
Kohorte 4															PISA

Bis 2012 beteiligte sich die Schweiz jeweils mit sprachregional und zum Teil kantonal repräsentativen Sichproben an PISA. Seit der Einführung der ÜGK bzw. seit 2015 beteiligt sich die Schweiz nur noch mit einer nationalen Stichprobe an PISA.

Mit dem Verzicht auf die ÜGK fallen die Überprüfungen der Bildungsziele bzw. des Erreichens der Grundkompetenzen weg. Trendstudien sind nur noch in Bezug auf das Ende der Schulzeit möglich. Da Eingangsmessungen fehlen, können die Leistungen der obligatorischen Schule im Sinne des Bildungsmonitorings nicht evaluiert werden.

Globale Einschätzung: Die Variante D führt die ÜGK nicht weiter und setzt damit Artikel 10 des HarmoS-Konkordats in Bezug auf die Bildungsstandards (Grundkompetenzen) nicht um. Sie fokussiert ausschliesslich auf die Ergebnisse der PISA-Studie und die dank den kantonal repräsentativen PISA-Zusatzstichproben möglichen Vergleiche im 11. Schuljahr.

Tabelle 5: Beurteilung der Variante D nach einzelnen Kriterien

Kriterium	**Beurteilung**
HarmoS-Konkordat 2007: Überprüfung des Erreichens der Bildungsziele, die der Harmonisierung dienen.	– Keine Überprüfung der Bildungsziele (Grundkompetenzen)
HarmoS-Konkordat 2007: Beitrag zu einer systematischen und kontinuierlichen Evaluation der Entwicklungen und Leistungen des Systems	± Trendmessungen zu Entwicklungen nur im 11. Schuljahr aufgrund der PISA-Tests – Keine Evaluation der Leistungen der Zyklen der obligatorischen Schule
Umsetzungsbeschluss 2007: ÜGK für interkantonalen Vergleich, PISA mit nationaler Stichprobe	– Nein, Rückkehr zum Zustand von vor 2015/16
Nutzen für - Bildungsmonitoring - Bildungssteuerung - Bildungsforschung	± Schaffung einer im Vergleich zu heute erweiterten Datengrundlage (Kompetenzen und Kontext) zum 11. Schuljahr aufgrund der der PISA-Studie. • regelmässige, wenn auch eher seltene Datenlieferungen • keine Kohortenstudien • nur marginale Möglichkeit, bei PISA eigene Interessen in Test und/oder Kontextualisierung einzubringen • mit anderen Datenbeständen der öffentlichen Hand verknüpfbar
Aufbau und Erhalt von Kompetenzen und Strukturen im Bereich Large-Scale-Assessments	– Kaum, nur Durchführung und Berichterstattung ± Lokale Forschung auf Grundlage der PISA-Daten möglich
Belastung der Schulen und Verwaltungen im Zusammenhang mit den Erhebungen	Innerhalb von 15 Jahren gibt es: 4 PISA-Haupterhebungen mit Stichproben, die jeweils umfangreicher sind als ÜGK-Stichproben im 11. Schuljahr, da kombiniert mit einer nationalen Stichprobe der 15-Jährigen; dazu 4 Piloterhebungen (*field trials*)
Durchschnittliche Kosten pro Jahr über 15 Jahre ab der ersten Haupterhebung	PISA zusätzlich ca. CHF 0.5 Mio. pro Jahr für kantonal repräsentative Stichproben; (Kosten für nationale Stichprobe 50% vom Bund getragen; Anteil EDK ca. 0.6 Mio. pro Jahr)

3.1.2 Erweiterung der ÜGK-Tests auf das ganze Leistungsspektrum

Die ÜGK erlaubt aktuell aufgrund des Fokus auf den Grundkompetenzen eine Aussage über den Erreichungsgrad der Grundkompetenzen, aber wenig bis keine Aussagen zu den Leistungsunterschieden zwischen den Schülerinnen und Schüler im mittleren und oberen Leistungsbereich. Damit eignen sich die Ergebnisse für bestimmte wichtige Analysen für eine evidenzgestützte Bildungssteuerung nicht (z.B. Analysen zur Entwicklung der Leistungen bestimmter Gruppen von Schülerinnen und Schülern im Vergleich zur Leistungsentwicklung anderer Gruppen innerhalb eines Zyklus).

Der Kosta HarmoS als Steuergremium der ÜGK ist zum Schluss gekommen, dass die ÜGK-Tests künftig nicht nur zur Überprüfung des Erreichens der Grundkompetenzen genutzt werden sollen, sondern gleichzeitig auch zur Erfassung der Kompetenzen auf dem ganzen Leistungsspektrum. Damit kann das Verhältnis von Aufwand und Nutzen deutlich verbessert werden. Zudem kann das Versprechen eingelöst werden, mit der ÜGK die früher üblichen Regions- und Kantonsvergleiche im Rahmen der PISA-Studie vollwertig zu ersetzen. PISA nutzt mehrstufige Kompetenzskalen. In der Berichterstattung spielt die Verteilung auf die Niveaus eine wichtige Rolle. In der Berichterstattung zur ÜGK soll weiterhin das Erreichen der Grundkompetenzen im Mittelpunkt stehen. Gleichzeitig sollen aber auch die Erkenntnisse zur Kompetenzverteilung für Aussagen genutzt werden.

Die Erweiterung der ÜGK-Tests auf das ganze Leistungsspektrum ist für alle drei ÜGK-Varianten (A-C) vorgesehen. Zusätzliche Bildungsziele werden dafür nicht verabschiedet. Die Umsetzung geschieht im Rahmen der vorgesehenen Entwicklung der Testframeworks sowie der Aufgaben und Tests. Sie führt nicht zu Entwicklungskosten, die über den veranschlagten Rahmen hinausgehen.

3.2 Anpassungen im Bereich der Governance

3.2.1 Anbindung der Aufgabendatenbank an eine wissenschaftliche Institution und Verzicht auf den Aufbau einer Aufgabendatenbank für andere Zwecke als die ÜGK

Die Aufgabendatenbank (ADB) ist Teil des GS EDK und ihre Geschäftsstelle ist bei der SKBF «domiziliert».

Gemäss Art. 3 des ADB-Reglements vom 23. Januar 2014 bestehen die Aufgaben der ADB insbesondere im Folgenden:

a. eine multifunktionale und mehrsprachige technische Struktur bereitzustellen,
b. eine gesamtschweizerische Messskala sicherzustellen,
c. einen Aufgaben- und Testentwicklungsprozess zu etablieren, in dessen Rahmen Aufgaben sowie Tests entwickelt und an den Grundkompetenzen geeicht werden können,
d. den Austausch von Aufgaben für die Überprüfung der Erreichung der nationalen Bildungsziele (Grundkompetenzen) sowie den Austausch von Aufgaben für Projekte der Kantone zu koordinieren.

Laut Art. 4 des ADB-Reglements (Dienstleistungen) kann die ADB zudem Aufträge zur Aufgabenentwicklung, für das Zurverfügungstellen von Aufgaben oder zur Eichung von Aufgaben an den Grundkompetenzen entgegennehmen.

Bis heute hat die ADB lediglich die Aufgaben wahrgenommen, die mit den bisherigen und den beschlossenen ÜGK-Zyklen im Zusammenhang stehen: Aufgaben- und Testentwicklung sowie Schwellenwertsetzung (vgl. oben Aufgabenbereiche b und vor allem c). Die Struktur zur Verwaltung und Auslieferung von Tests und Aufgaben wurde letztlich anderweitig umgesetzt. Die Angebote an die Regionen und Kantone (vgl. Art.3 und Art.4) wurden nicht in Anspruch genommen.

Das wissenschaftliche Gutachten zur ÜGK empfiehlt, die ADB in die existierenden wissenschaftlichen Strukturen zur Umsetzung der ÜGK zu integrieren, weil sie wissenschaftliche Arbeiten ausführt und das

Knowhow nachhaltig gesichert werden muss und weil es zahlreiche Schnittstellen zu anderen Umsetzungsarbeiten gibt, die möglichst effizient bewirtschaftet werden sollen.

Der Kosta HarmoS befürwortet aus den genannten Gründen eine Integration der ADB in die wissenschaftlichen Strukturen, die vom Interfaculty Centre for Educational Research (ICER) der Universität Bern koordiniert bzw. geleitet werden.

Gleichzeitig schlägt er vor, die Aufgaben gemäss Reglement, die nicht direkt mit der ÜGK zu tun haben und welche die ADB bisher nicht wahrgenommen hat, ersatzlos zu streichen.

3.2.2 Verantwortung für den Kontextfragebogen und die Schwellenwerte beim wissenschaftlichen Konsortium

Der Plenarversammlung obliegt gemäss ÜGK-Reglement vom 8. Mai 2014 die Genehmigung des Schwellenwertes («Grundkompetenzen erreicht») auf den Testskalen, den die ADB zuvor mit wissenschaftlichen Methoden ermittelt hat.

Der EDK-Vorstand genehmigt gemäss aktuellen Rechtsgrundlagen den Inhalt des ÜGK-Kontext-fragebogen für die Schülerinnen und Schüler. Diese Praxis hat sich etabliert, weil vor allem die Fragen zum sozialen Hintergrund oft als heikel empfunden wurden. Die Genehmigung kurz vor den Erhebungen wurde allerdings von verschiedenen Akteuren als unbefriedigend empfunden.

Das wissenschaftliche Gutachten erachtet Schwellenwertsetzung und Fragebogenentwicklung als rein wissenschaftliche Aufgabe und äussert die Befürchtung, dass eine Beteiligung der politischen Instanzen an diesen Arbeiten und die Möglichkeit der Einflussnahme der Glaubwürdigkeit der Ergebnisse schaden. Das Gutachten schlägt deshalb vor, wissenschaftliche Aufgaben der Wissenschaft zu überlassen und dafür den Auftrag von Bildungspolitik und Bildungsverwaltung an die wissenschaftlichen Institutionen, die die ÜGK umsetzen, konkreter auszuarbeiten als bisher und gemeinsam zu verabschieden. Dafür eignet sich ein sogenanntes Framework, ein konzeptioneller Rahmen für die Umsetzung der ÜGK, insbesondere für die Bereiche Testen und Kontextualisierung (Fragebogen). Die PISA-Studie arbeitet mit Test- und Fragebogenframeworks. Vor jedem neuen Zyklus werden sie in Zusammenarbeit mit externen Fachexpertinnen und -experten aktualisiert, den auftraggebenden Ländern zur Begutachtung vorgelegt und schliesslich verabschiedet. Sie bilden dann die Grundlage für die Auftragsvergabe und die Umsetzung der Studie.

Der Kosta HarmoS unterstützt die vorgeschlagene Trennung von bildungspolitischen und wissenschaftlichen Aufgaben sowie die Definition der Schnittstelle zwischen Auftraggeber und Auftragnehmer mithilfe von Frameworks.

3.2.3 Stärkung der Stellung der Fachdidaktiken

In den Nullerjahren waren Fachdidaktikerinnen und Fachdidaktiker im HarmoS-Projekt intensiv an der Entwicklung von Kompetenzmodellen und der Beschreibung von Bildungsstandards für die Fachbereiche Schulsprache, Mathematik, Fremdsprachen und Naturwissenschaften beteiligt. In der ÜGK spielen die Fachdidaktiken vor allem in der Aufgabenentwicklung eine wichtige Rolle, sind aber lediglich über die ADB in die Projektstruktur eingebunden und spielen als solche in den Projektstrukturen keine Rolle. Gemäss dem wissenschaftlichen Gutachten soll die Perspektive der Fachdidaktiken (FD) der in der ÜGK getesteten Fachbereiche in der ÜGK systematisch berücksichtigt werden, einerseits wegen ihrer fachspezifischen Perspektive auf Anlage und Nutzung der ÜGK und andererseits wegen ihrer Nähe zur Aus- und Weiterbildung der Lehrpersonen.

Der Kosta HarmoS teilt das Anliegen des Gutachtens weitgehend und schlägt zwei Massnahmen vor, um die Perspektive der Fachdidaktik in der ÜGK besser zur Geltung zu bringen: 1) einen direkten Einbezug der Fachdidaktik in die Arbeitsgruppe Testentwicklung der Projektleitung ÜGK; 2) eine Vertretung der Fachdidaktik in der vorgeschlagenen Expertengruppe der EDK zur Begleitung der ÜGK (s. nächster Abschnitt). Zudem erachtet er es als wünschenswert, insbesondere auch Fachdidaktikerinnen und Fach-didaktiker mit Unterrichtserfahrung bei den getesteten Schülerinnen und Schülern in die Aufgabenentwick-lung einzubinden, um den Praxisbezug sicherzustellen.

3.2.4 Einsetzung einer interregionalen, internationalen und interdisziplinären Expertengruppe

Ein Aufbaugremium, bestehend aus Vertreterinnen und Vertretern der Universität Bern und Genf sowie der Scuola universitaria professionale della Svizzera italiana (SUPSI) haben bis Ende 2016 im Auftrag der EDK den Entwurf einer Geschäftsordnung für ein wissenschaftlichen Konsortium Bildungsforschung und -evaluation (WIKO BFE) ausgearbeitet. Dieser Entwurf befindet sich im Anhang zu den Kooperationsvereinbarungen mit der Universität Bern betreffend die Realisierung der aktuellen ÜGK- und PISA-Zyklen. Die dort skizzierte Verankerung der Large-Scale-Assessments in der Hochschullandschaft der Schweiz sowie der Zugang zu höchster Expertise wurden nie wie vorgesehen umgesetzt. Vorgesehen waren eine Professur für Bildungsforschung und -evaluation an der Universität Bern, ein wissenschaftlicher Beirat mit Beratungs-, Leitungs- und Entscheidungsfunktionen (bestehend aus Vertretungen der Universitäten Bern und Genf sowie der SUPSI und evtl. weiteren), eine Delegiertenversammlung von fachlich interessierten Vertreterinnen und Vertretern von Schweizer Hochschulen sowie ein internationaler Rat von drei Expertinnen und Experten aus verschiedenen Sprachräumen.

Als teilweisen Ersatz dafür, schlägt das GS EDK vor, eine interregionale, internationale und interdisziplinäre Gruppe von Expertinnen und Experten einzusetzen, die das ÜGK-Projekt begleitet und dabei spezifisch auch die Interessen der EDK bei der Entwicklung, Verabschiedung und Umsetzung der Frameworks vertritt. Bei der Zusammensetzung wird auf fachliche Breite und Expertise sowie auf die Vertretung der Regionen geachtet. Für die Beschäftigung mit den Frameworks müssen unter anderem auch die Fachdidaktiken berücksichtigt werden, die zurzeit nur für bei der Aufgabenentwicklung mitwirken.

3.2.5 Zusammenarbeit des wissenschaftlichen Konsortiums bzw. der Universität Bern mit den regionalen Durchführungszentren

Aufgrund des Organisationsreglements über die Durchführung der ÜGK ist die Universität Bern bei der ÜGK zur Zusammenarbeit mit sprachregionalen Institutionen verpflichtet. Das Reglement legt auch die dezentrale Durchführung der Erhebungen fest. Wie der vorige Abschnitt zeigt, ist die sprachregionale und dezentrale Verankerung des ÜGK-Projekts weniger umfassend umgesetzt, als insbesondere im Entwurf der Geschäftsordnung für das WIKO BFE vorgesehen. Die regionalen Durchführungszentren an PHSG, SRED und SUPSI sind dagegen intensiv in die Projektleitung und in die institutionsübergreifenden Arbeitsgruppen des ICER zur Umsetzung der ÜGK eingebunden. Die drei Durchführungszentren unterhalten auch die Kontakte zu den Kantonen und leiten die Datenerhebungen in den Schulen.

Wenn die Leitungsfunktion des ICER gestärkt und das Projekt stärker zentralisiert würde, könnten voraussichtlich die Kosten für ÜGK gesenkt werden. Es könnten mehr inhaltliche Aufgaben zentral durch das ICER anstatt durch die institutionsübergreifenden Arbeitsgruppen (z.B. AG Kontextfragebogen, AG Testentwicklung, AG Testauslieferung, AG Erhebungsmanagement) wahrgenommen werden. Auch das Netzwerk der kantonalen Referenzpersonen könnte zentral bedient werden. Falls in Zukunft die ÜGK-Tests online ausgeliefert werden können, könnten die Erhebungen in den Schulen eventuell von privaten Auftragnehmern übernommen werden.

Wahrscheinlichen Kosteneinsparungen bei einer Zentralisierung steht eine Schwächung der sprachregionalen Verankerung der ÜGK gegenüber. Das Knowhow im Bereich der Leistungsmessung dürfte abseits des Zentrums dürfte mit der Zeit ebenfalls schwinden. Aufgrund der bisherigen positiven Erfahrungen mit den aktuellen Strukturen möchte das ICER selbst auf eine konsequente Zentralisierung verzichten.

Um die sprachregionale Abstützung der Arbeiten nicht zu gefährden, schlägt der Kosta HarmoS vor, die bewährte Zusammenarbeit vorderhand beizubehalten und im Laufe der kommenden Jahre gemeinsam mit der Universität Bern Anpassungen zu prüfen.

3.3 Kostenfolgen

Wie in Abschnitt 1.4.3 gezeigt, werden aktuell pro Jahr (ohne coronabedingte Zusatzkosten) für die ÜGK 1,13 Mio. CHF und für die ADB 0,37 Mio. CHF an Kantonsbeiträgen geleistet. Mit den Beiträgen für sechs Jahre, d.h. mit rund 9 Mio. CHF können zwei Erhebungen vollständig finanziert werden.

Aus den Tabellen 1 bis 5 in Abschnitt 3.1 geht hervor, dass die vorgeschlagene Neugestaltung und Aufwertung der ÜGK jährliche Mehrkosten zur Folge hat, die v.a. mit der grösseren Zahl an Erhebungen im Zusammenhang stehen. Die angegebenen Kosten von 1.95 Mio. CHF (Variante C), 2.25 Mio. CHF (Variante A) und 2.8 Mio. CHF pro Jahr (Variante B) zeigen dies deutlich. Bei Variante D wird die ÜGK aufgegeben. Im Zusammenhang mit der Durchführung der PISA-Erhebungen bei einer Stichprobe des 11. Schuljahres, die umfangsmässig etwa einer ÜGK-Stichprobe auf dieser Stufe entspricht, entstehen gegenüber heute Mehrkosten von rund 0.5 Mio. pro Jahr. Die Teilnahme an PISA ist wesentlich billiger als die Durchführung von ÜGK-Erhebungen, weil praktisch keine eigenen Entwicklungsarbeiten geleistet werden müssen.

Quellenverzeichnis

Ambühl, H. (2011). Die interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14.6.2007: Kommentar, Entstehungsgeschichte und Ausblick, Instrument. EDK. http://edudoc.ch/record/96777/files/harmos-konkordat_d.pdf

Angelone, D., & Keller, F. (2019a). Überprüfung des Erreichens der Grundkompetenzen (ÜGK) im Fach Mathematik im 11. Schuljahr: Technische Dokumentation zur Testentwicklung und Skalierung. Aufgabendatenbank der EDK. http://uegk-schweiz.ch/wp-content/uploads/2019/05/%C3%9CGK2016_Technischer-Bericht_ADB.pdf

Angelone, D., & Keller, F. (2019b). Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in den Fächern Schulsprache und erste Fremdsprache im 8. Schuljahr: Technische Dokumentation zur Testentwicklung und Skalierung. Aufgabendatenbank der EDK. http://www.vecof-svizzera.ch/wp-content/uploads/2019/06/%C3%9CGK2017_Technischer-Bericht_ADB.pdf

Angelone, D., Keller, F., Arnold, J., Jaun-Holderegger, B., Hauser, M., Eckstein, B., Erzinger, A., Leutwiler, L., Nell-Tuor, N., & Schönenberger, S. (2019). Machbarkeitsstudie zur Überprüfung der Grundkompetenzen im 4. Schuljahr [Unveröffentlichter Bericht im Auftrag der EDK]. Aufgabendatenbank der EDK.

Buschor, E., Gilomen, H., & McCluskey, H. (2003). PISA 2000: Synthese und Empfehlungen. BFS; EDK. http://edudoc.ch/record/24195/files/8d.pdf

Crotta, F., Salvisberg, M., & Zanolla, G. (2020). Note scolastiche e competenze fondamentali in matematica degli allievi dell'11° anno HarmoS in Ticino. Swiss Journal of Educational Research, 42(2), 323–342. https://doi.org/10.24452/sjer.42.2.3

EDK. (1970). Konkordat über die Schulkoordination vom 29. Oktober 1970. https://edudoc.ch/record/1987/files/1-1d.pdf

EDK. (2001a). Leitlinien der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren (EDK) vom 5.7.2001. https://edudoc.ch/record/24715

EDK. (2001b). Tätigkeitsprogramm der EDK ab 2001 nach Arbeitschwerpunkte vom 7.6.2001. http://edudoc.ch/record/112626/files/TP2001d.pdf

EDK. (2004a). Sprachenunterricht in der obligatorischen Schule: Strategie der EDK und Arbeitsplan für die gesamtschweizerische Koordination: Beschluss vom 25. März 2004. http://edudoc.ch/record/30008/files/Sprachen_d.pdf

EDK. (2004b). HarmoS: Zielsetzungen und Konzeption. https://edudoc.ch/record/33485?ln=de

EDK. (2007a). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007. https://edudoc.ch/record/24711

EDK. (2007b). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat); Umsetzung auf der Ebene der interkantonalen Koordination vom 25./26. Oktober 2007. http://edudoc.ch/getfile.py?recid=30023

EDK. (2009). Beteiligung an PISA 2012 (Schwerpunkt Mathematik): Beschluss Plenarversammlung vom 29./30.10.2009. http://edudoc.ch/record/102486/files/05_d.pdf

EDK. (2011a). Grundkompetenzen für die Fremdsprachen: Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16.6.2011. http://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf

EDK. (2011b). Grundkompetenzen für die Mathematik: Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16.6.2011. http://edudoc.ch/record/96784/files/grundkomp_math_d.pdf

EDK. (2011c). Grundkompetenzen für die Naturwissenschaften: Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16.6.2011. http://edudoc.ch/record/96787/files/grundkomp_nawi_d.pdf

EDK. (2011d). Grundkompetenzen für die Schulsprache: Nationale Bildungsstandards. Frei gegeben von der EDK-Plenarversammlung am 16.6.2011. http://edudoc.ch/record/96791/files/grundkomp_schulsprache_d.pdf

EDK. (2012a). Teilnahme der Schweiz an PISA 2015: Beschlussfassung: Beschluss vom 25.10.2012. https://edudoc.ch/record/105012/files/PB_pisa_d.pdf

EDK. (2012b). Überprüfung der Erreichung der Grundkompetenzen: Konzept und Finanzierung der Aufgabendatenbank: Beschlussfassung: Beschluss vom 25.10.2012. http://edudoc.ch/record/105010/files/PB_aufgabendatenbank_d.pdf

EDK. (2013a). Überprüfung der Erreichung der Grundkompetenzen; Konzept: Verabschiedung: Beschluss vom 20.6.2013. https://edudoc.ch/record/107770/files/PB_grundkompetenzen_d.pdf

EDK. (2013b). Reglement des Koordinationsstabes für die Umsetzung der Interkantonalen Vereinbarung über die Harmonisierung der obligatorischen Schule (Kosta HarmoS) vom 24. Januar 2008. http://edudoc.ch/record/38020/files/KostaHarmos_d.pdf

EDK. (2014a). Organisationsreglement für die Aufgabendatenbank der Schweizerischen Konferenz der kantonalen Erziehungsdirektoren vom 23. Januar 2014. http://edudoc.ch/record/110605/files/fj_Organisationsreglement_d_140123.pdf

EDK. (2014b). Organisationsreglement über die Durchführung der Überprüfung des Erreichens der Grundkompetenzen vom 8. Mai 2014. http://edudoc.ch/record/113128/files/Organisationsregl_UeGK_d.pdf

EDK. (2017). Überprüfung des Erreichens der Grundkompetenzen: Planung zu weiteren ÜGK-Erhebungen ab 2020: Beschluss Plenarversammlung vom 22. Juni 2017. http://edudoc.ch/record/127425/files/pb_grundkompetenzen_erhebungen_2020_d.pdf

EDK. (2018a). Überprüfung des Erreichens der Grundkompetenzen: Datennutzungskonzept: Verabschiedung: Beschluss Plenarversammlung vom 22. März 2018. http://edudoc.ch/record/130620/files/pb_datennutzungskonezpt_d.pdf

EDK. (2018b). Überprüfung des Erreichens der Grundkompetenzen; zu testende Fachbereiche in der Erhebung 2020, 11. Schuljahr: Beschlussfassung: Beschluss Plenarversammlung vom 26. Oktober 2018. http://edudoc.ch/record/133108/files/pb_fachbereiche_d.pdf

EDK. (2019a). Überprüfung des Erreichens der Grundkompetenzen: Bericht zu den Erhebungen 2016 und 2017. EDK. https://edudoc.ch/record/204069

EDK. (2019b). Grundkompetenzen Sprachen und Mathematik: Ergebnisse der Erhebungen 2016 und 2017. Kurzinformation 24.5.2019. https://edudoc.ch/record/217180/files/kurzinfo-uegk-d.pdf

EDK. (2019c). Überprüfung des Erreichens der Grundkompetenzen; zu testende Fachbereiche in der Erhebung 2022: Beschluss: Beschluss Plenarversammlung vom 25. Oktober 2019. http://edudoc.ch/record/206945/files/PB_fachbereiche_erhebung_2022_d.pdf

EDK. (2020a). Überprüfung des Erreichens der Grundkompetenzen; Verzicht auf die Durchführung der Erhebung 2020 und Erarbeitung eines neuen Zeitplans für die Erhebungen 2020 und 2022: Beschluss: Korrespondenzbeschluss Plenarversammlung vom 1. April 2020. EDK. http://edudoc.ch/record/209009/files/PLE-2020-2_verzicht-durchfuehrung-uegk-2020.pdf

EDK. (2020b). Überprüfung des Erreichens der Grundkompetenzen; Angepasster Zeitplan zur Durchführung der Erhebungen im 4. Und 11. Schuljahr: Beschluss der Plenarversammlung vom 25. Juni 2020. http://edudoc.ch/record/210793/files/PV_Zeitplan_UEGK.pdf

EDK. (2021a). Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat) vom 14. Juni 2007: Stand kantonale Beitrittsverfahren. EDK. https://edudoc.ch/record/210894/files/liste_rat_df.pdf

EDK. (2021b). Datennutzung im Rahmen der Überprüfung des Erreichens der Grundkompetenzen: Konzept. https://edudoc.ch/record/218327/files/VB070521-konzept-uegk-d.pdf

EDK/IDES. (2021). Kantonsumfrage: Stand: Schuljahr 2020-2021: Standardisierte Leistungstests / Primarstufe (Jahre 3-8) / Sekundarstufe I: Frage 121: Werden gemäss kantonalen Regelungen für alle Schülerinnen und Schüler obligatorische standardisierte Schulleistungstests (z.B. Stellwerk, Checks, Orientierungsarbeiten, Abschlussprüfungen) eingesetzt? https://edudoc.ch/record/212326/files/121.pdf

Erzinger, A., Hauser, M., Dutrevis, M., Hascher, T., Keller, R., Lenz, P., & Soussis, A. (2019). Erläuterungen zu den Skalen des Kontextfragebogens der ÜGK Sprachen 2017: Theoretischer Hintergrund, Inhalte und Konstrukte. Universität Bern, Päd. Hochschule St. Gallen, Service de la recherche en éducation (SRED), Universität Fribourg. http://uegk-schweiz.ch/wp-content/uploads/2019/09/%C3%9CGK2017_Sprachen_KFB_def.pdf

Fischbach, A., & Ugen, S. (2018). ÜGK/COFO Mathematics 2016: Audit report: Commissioned by EDK/CDIP upon the request of KOSTA HarmoS. [s.n.]. http://edudoc.ch/record/204067/files/UeGK_Audit_Report_230218.pdf

IRDP. (2021). EPROCOM - banque d'items. https://www.irdp.ch/institut/eprocom-banque-items-3673/irdp-fromsearch-1.html

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2004). Zur Entwicklung nationaler Bildungsstandards – Expertise (Bd. 1). Bundesministerium für Bildung und Forschung (BMBF). https://edudoc.ch/record/33468/files/develop_standards_nat_form_d.pdf

Konsortium HarmoS Fremdsprachen, Schneider, G., Lenz, P., & Studer, T. (2009). Fremdsprachen: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Standards). Stand: 18. Juni 2009. http://edudoc.ch/record/87025/files/L2_wissB_25_1_10_d.pdf

Konsortium HarmoS Naturwissenschaften+. (2010). Naturwissenschaften: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Standards): Stand Juli 2009, mit Ergänzungen und Korrekturen Januar 2010. Konsortium HarmoS Naturwissenschaften+. http://edudoc.ch/record/86401/files/harmoS_kurzbericht_neu.pdf

Konsortium HarmoS Schulsprache. (2010). Schulsprache: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Standards): Stand: 17. Januar 2010. [s.n]. http://edudoc.ch/record/87022/files/L1_wissB_25_1_10_d.pdf

Konsortium Mathematik, Linneweber-Lammerskitten, H., Wälti, B., & Moser Opitz, E. (2009). HarmoS Mathematik: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Basisstandards): Stand: 13. Dezember 2009. [s.n]. http://edudoc.ch/record/87030/files/math_kurzbericht_2009_d.pdf

Nidegger, C., Roos, E., Petrucci, F., Verner, M., Helbling, L. A., Arnold, E., Erzinger, A. B., Pham, G., Ambrosetti, A., Crotta, F., Salvisberg, M., Moretti, A., Angelone, D., Keller, F., Leutwiler, L., Lindauer, T., Lenz, P., Aeppli, T., Karges, K., … Epars, F. (2019). Überprüfung der Grundkompetenzen: Nationaler Bericht der ÜGK 2017: Sprachen 8. Schuljahr. EDK; SRED. http://edudoc.ch/record/204072/files/UGK_2017_Completo_DE.pdf

Nidegger, C., Roos, E., Petrucci, F., Verner, M., Helbling, L. A., Arnold, E., Erzinger, A. B., Pham, G., Schönenberger, S., Crotta, F., Ambrosetti, A., Salvisberg, M., Angelone, D., Keller, F., & Linneweber-Lammerskitten, H. (2019). Überprüfung der Grundkompetenzen: Nationaler Bericht der ÜGK 2016:

Mathematik 11. Schuljahr. EDK: SRED. http://edudoc.ch/record/204075/files/UGK_2016_Completo_DE.pdf

Petrucci, F., & Roos, E. (2020). Atteinte des compétences fondamentales dans le canton de Genève: Que nous enseignent les enquêtes COFO 2016 et 2017 ? SRED. https://edudoc.ch/record/213034

Ramseier, E., Moser, U., Moreau, J., & Antonietti, J.-P. (2008). Schlussbericht der HarmoS-Methodologiegruppe. EDK. http://www.lehrplanforschung.ch/wp-content/uploads/2011/06/Methodologie-Schlussbericht-_HarmoS11.pdf

Roth, M., Ruf, I., Sánchez Abchi, V., Soussi, A., Weiss, L., & Marc, V. (2021). EpRoCom: Dispositif romand de mutualisation de tâches évaluatives (IRDP Focus). https://www.irdp.ch/data/secure/3586/document/2021_irdp_focus_eprocom.pdf

SKBF CSRE. (2018). Bildungsbericht. http://www.skbf-csre.ch/bildungsbericht/bildungsbericht/

WBF, & EDK. (2019). Chancen optimal nutzen: Erklärung 2019 zu den gemeinsamen bildungspolitischen Zielen für den Bildungsraum Schweiz. http://edudoc.ch/record/205824/files/erklaerung_20190627_d.pdf

Anhang 1: Weitere Aspekte der Umsetzung der ÜGK

Testentwicklung, Schwellenwertsetzung

Wie in Abschnitt 1.3.3 beschrieben, macht bereits der Auditbericht zur ÜGK 2016 (Fischbach & Ugen, 2018) Vorschläge zur Stärkung der Aufgaben- und Testentwicklung. Diese werden vor allem im Zusammenhang mit den Arbeiten an der ÜGK 2024 umgesetzt. Das Gutachten zur ÜGK würdigt die bereits getroffenen Massnahmen, identifiziert aber nach wie vor Problemstellen und macht dazu Empfehlungen. Eine Reihe von Aspekten wird im Folgenden dargestellt und diskutiert. Zudem weist der Kosta HarmoS auf die Änderungen hin, die nötig werden, wenn die ÜGK-Tests in Zukunft nicht nur das Erreichen der Grundkompetenzen überprüfen, sondern Kompetenzen über das ganze Kompetenzspektrum hinweg erfassen wollen.

Schwächen im Bereich des Test-Frameworks und der Grundlagen für die Testentwicklung

Fachdidaktische Verankerung: Das Gutachten stellt fest, dass bei der ÜGK ausgebaute Test-Frameworks, wie sie beispielsweise von der PISA-Studie her bekannt sind, fehlen. Gemäss Gutachten beschreiben solche Frameworks insbesondere auch aus fachdidaktischer Perspektive, was unter den Kompetenzen verstanden wird, die getestet werden sollen. Aufgabenbeispiele dienen der Illustration und dem besseren Verständnis. Mithilfe der Aufgaben und der Beschreibung ihrer Merkmale können die Anforderungen an Grundkompetenzen verständlich gemacht und diskutiert werden. Dies bringt Vorteile für die Verankerung der Bildungsziele in der Fachdidaktik, in der Ausbildung von Lehrpersonen und in der Unterrichtspraxis. Aus Sicht der Gutachtengruppe sollte jeder ÜGK-Test mit einem jeweils aktualisierten Framework begründet werden. Die Testframeworks, wie sie ansatzweise in der Berichterstattung zur ÜGK 2017 vorhanden sind, sollten ausgebaut und die Kompetenzbereiche in einer Art beschrieben werden, die für die fachdidaktische Ausbildung und letztlich die Unterrichtspraxis besser nutzbar ist.

Konkrete Grundlagen für die Testentwicklung: Die freigegebenen Bildungsziele für die vier Fachbereiche beschreiben im Wesentlichen, was die Schülerinnen und Schüler mindestens können sollen, benennen also Art und Inhalt der Anforderungen. Dies genügt aber erfahrungsgemäss für sich allein nicht, um eine systematische Aufgabenentwicklung zu gewährleisten, weil in den Beschreibungen wichtige Anhaltspunkte für die konkrete Umsetzung in Testaufgaben fehlen. Um gesicherte Grundlagen für die Aufgabenentwicklung zu haben und die Aufgabenschwierigkeit von vornherein besser steuern zu können, schlägt das Gutachten vor, bereits eingesetzte Aufgaben mit bekannten statistischen Eigenschaften daraufhin zu analysieren, wie sich deren Merkmale auf die Schwierigkeit auswirken. Weiter sollen Merkmale identifiziert werden, die sich in verschiedenen Lehrplan- und/oder Sprachregionen unterschiedlich auf die Aufgabenschwierigkeit auswirken.

Nach Einschätzung des Kosta HarmoS sind die Vorschläge des Gutachtens zur Stärkung der fachdidaktischen Verankerung der ÜGK und zur Nutzung von Test- und Aufgabenanalysen für eine gezielte Aufgabenentwicklung relevant und zielführend. Dadurch, dass zukünftige ÜGK-Tests a) das Erreichen der Grundkompetenzen überprüfen, b) Kompetenzen über möglichst das ganze vorhandene Kompetenzspektrum hinweg testen sollen (vgl. vor allem 2.1.4), und weil zudem ein Längsschnittdesign umgesetzt werden soll (vgl. Abschnitt 2.2), erweitert sich der Auftrag an die Fachdidaktiken (und an die Psychometrie):

- Die Kompetenzmodelle der Grundkompetenzen müssen um eine vertikale Dimension ergänzt werden durch a) eine stufenübergreifende Charakterisierung der Kompetenzentwicklung und b) Skalen, die das Leistungsspektrum pro Testzeitpunkt beschreiben, einschliesslich Aufgabenbeispielen.

– Es müssen gesicherte konzeptionelle Grundlagen für die Entwicklung von Tests[7] geschaffen werden, die sowohl im Bereich der Grundkompetenzen als auch über ein breites Kompetenzspektrum hinweg zuverlässig testen. Für den oberen Skalenbereich gibt es aus den bisherigen ÜGK-Zyklen nicht viele bereits eingesetzte Aufgaben, die analysiert werden könnten, um Schwierigkeitsmerkmale zu gewinnen. Die Frameworks für die Testentwicklung werden einen Vorschlag enthalten müssen, auf welche Weise Schülerinnen und Schüler mit weit auseinanderliegenden Kompetenzen effizient und gleichzeitig präzise getestet werden können. Beispielsweise kann ein *multi-stage adaptive testing* (MSAT) wie bei PISA umgesetzt werden. Beim MSAT wird die Schwierigkeit der Aufgabenblöcke ab dem zweiten Block aufgrund der früher erreichten Ergebnisse individuell festgelegt.

Passung der Aufgabenschwierigkeit

Die ÜGK-Tests haben bisher bekanntlich vor allem den Zweck, das Erreichen der Grundkompetenzen zu überprüfen. Was das Erreichen der Grundkompetenzen in Bezug auf die Testergebnisse bedeutet, entscheidet sich erst bei der Schwellenwertsetzung definitiv. Beim Bookmark-Standardsetting, das bei der ÜGK verwendet wird, erhält ein gemischtes Gremium von Fachleuten ein Heft, in dem die einzelnen Items nach aufsteigender Schwierigkeit (gemäss Testergebnis) angeordnet sind. Die Aufgabe besteht darin zu entscheiden, welche Aufgabe Schülerinnen und Schüler, die gerade über Grundkompetenzen verfügen, noch mit einer Wahrscheinlichkeit von 62% lösen können. Aufgrund dieser Entscheidung kann anschliessend berechnet werden, welcher Anteil der Getesteten über Grundkompetenzen verfügt.

Wenn eine zentrale Funktion der ÜGK-Tests darin besteht, die Schülerinnen und Schüler über der Schwelle («Grundkompetenzen erreicht») von denen unter der Schwelle zu unterscheiden, dann sollten genügend Aufgaben bereitstehen, die sich dazu besonders eignen. Gemäss der Testtheorie vermögen Aufgaben ähnlich kompetente Testteilnehmende dann besonders gut voneinander zu unterscheiden, wenn diese Testteilnehmenden bei den betreffenden Aufgaben eine Lösungswahrscheinlichkeit von bloss etwa 50% haben. Dies trifft grundsätzlich auch für die Gruppe der Schülerinnen und Schüler zu, die ungefähr über Mindestkompetenzen verfügen. Für die Schwellenwertsetzung wiederum ist es wichtig, dass genügend Aufgaben vorhanden sind, die von den Schülerinnen und Schülern an der Schwelle mit einer Wahrscheinlichkeit von etwa 62% korrekt gelöst werden können. Einfachere Aufgaben, bei denen eine immer höhere Lösungswahrscheinlichkeit bis hin zu etwa 80% erreicht wird, sind nützlich, um illustrieren zu können, was mindestkompetente Schülerinnen und Schüler zunehmend sicherer können. Einfache Aufgaben spielen auch eine Rolle bezüglich der Testmotivation: Wenn zu viele Aufgaben als schwierig empfunden werden, besteht die Gefahr, dass die Leistungsbereitschaft rasch verloren geht. Aufgaben mit einer noch höheren Lösungswahrscheinlichkeit als 80% dienen in der ÜGK vor allem dem Zweck, die Kompetenzen von ganz schwachen Schülerinnen und Schülern voneinander zu unterscheiden. Umgekehrt helfen schwierigere Aufgaben, die Mindestkompetente nur noch mit geringer Wahrscheinlichkeit lösen können, bei der Differenzierung von stärkeren Schülerinnen und Schülern. Für Studien zu Erklärungsansätzen für Leistungsunterschiede und weitere Studien zu Zusammenhängen zwischen Testergebnissen und anderen Variablen ist es wichtig, dass auch bei den Stärksten und den Schwächsten noch eine Streuung besteht, also möglichst wenige Getestete das Punktemaximum oder -minimum erreichen.

Da unter normalen Umständen nur ein kleiner Anteil der Getesteten die Grundkompetenzen nicht erreichen sollte, ergibt sich aus den obigen Überlegungen, dass in einem Test zur Überprüfung des Erreichens der Grundkompetenzen ein Grossteil der Aufgaben für die meisten Schülerinnen und Schüler einfach bis sehr einfach sein sollte.

Die Gutachtengruppe kritisiert an den bisherigen ÜGK-Tests zu mehreren Kompetenzbereichen die ungenügende Passung der Aufgaben zum Testzweck. Zu viele Aufgaben sind zu schwierig und zu wenige

[7] In der Testforschung ist zum Teil von *constructor-oriented scales* (dt. aufgabenorientierten Skalen) die Rede, die für den Zweck einer gezielten Aufgabenkonstruktion entwickelt werden. Sie konkretisieren die Anforderungsmerkmale von Aufgaben, während Kompetenzskalen (oder user-oriented scales), die anhand der Anforderungen festgestellten Kompetenzen beschreiben.

einfach genug, um für die Überprüfung des Erreichens der Grundkompetenzen besonders nützlich zu sein. Insgesamt sieht sie dringenden Handlungsbedarf – dies umso mehr, als bei den Analysen und Befragungen zum Gutachten festgestellt wurde, dass unter den betreffenden Akteuren nicht Konsens darüber besteht, was das Messen des Erreichens der Grundkompetenzen bedeutet. Das Gutachten schlägt deshalb vor, dass sich Psychometrie und Fachdidaktik der verschiedenen Fachbereiche eine gemeinsame Position dazu erarbeiten, wie das Erreichen der Grundkompetenzen gemessen wird und wie man von der Beschreibung der Bildungsziele zu einer fachdidaktisch interpretierbaren Unterscheidung zwischen «Mindestkompetenzen erreicht» und «Mindestkompetenzen nicht erreicht» gelangt.

Der Kosta HarmoS stützt die Empfehlung des Gutachtens, dass sich Psychometrie und Fachdidaktiken gemeinsam ein Grundverständnis davon erarbeiten müssen, wie das Erreichen der Grundkompetenzen überprüft werden soll, und weist gleichzeitig darauf hin, dass es im Hinblick auf die Zukunft der ÜGK auch darum gehen muss, ein gemeinsames Grundverständnis für das Testen des ganzen Kompetenzspektrums zu entwickeln.

Aufgabenübersetzungen

Unter anderem aufgrund der Empfehlungen des Auditberichts zur ÜGK 2016 wird den Übersetzungen der Aufgaben inzwischen mehr Aufmerksamkeit geschenkt. Das Gutachten empfiehlt, in den Bemühungen noch weiter zu gehen und internationale Standards zum Übersetzen von Tests zu befolgen.

Der Kosta HarmoS hält ein Vorgehen nach internationalen Standards nicht für prioritär. Erfahrungsgemäss kann durch die sprachregionenübergreifende Zusammenarbeit bei der Testentwicklung sowie den gezielten Einbezug von geeigneten Institutionen des Bildungssystems bei einer sorgfältig geplanten Zusammenarbeit und wiederholten Überprüfungen eine hohe Qualität erreicht werden.

Studien zur Testvalidität

Das Gutachten weist darauf hin, dass (publizierte externe) Validierungsstudien zur ÜGK weitgehend fehlen. Es nennt insbesondere Studien, die überprüfen, inwieweit Ergebnisse bei den ÜGK-Tests mit den Ergebnissen bei anerkannten Tests oder schulischen Bewertungen zu denselben Kompetenzbereichen übereinstimmen (*concurrent validity*) und schlägt zudem repräsentative Befragungen von Lehrpersonen zur Passung der Tests zum Unterricht vor (*curricular validity*). Das Gutachten sieht solche Studien als vordringlich an und empfiehlt, zuerst Validierungsstudien durchzuführen und erst dann mit der weiteren Ausdifferenzierung der Testskalen zu beginnen.

Aus Sicht des Kosta HarmoS sind (veröffentlichte) Validierungsstudien zur Testvalidität sehr erwünscht. Die früher erwähnte Machbarkeitsstudie zur ÜGK im 4. Schuljahr (Angelone et al., 2019) ist beispielsweise nur als graues Papier zugänglich, und die Massnahmen zur Validierung der Testaufgaben während der Entwicklung werden im Rahmen der umfassenden Berichterstattung zu den ÜGK-Zyklen nur kurz dokumentiert. Aus der Westschweiz und dem Tessin liegen erfreulicherweise erste Studien zum Verhältnis der ÜGK-Ergebnisse zu den Schulnoten der Schülerinnen und Schüler vor (Crotta et al., 2020; Petrucci & Roos, 2020). In Anbetracht der teilweisen Neuausrichtung der ÜGK hält der Kosta HarmoS das vorgeschlagene Nacheinander von Validierungsstudien und weiteren Arbeiten zur Ausdifferenzierung der Testskalen nicht für opportun. Es werden auch Validierungsstudien notwendig sein, die erst nach Vorliegen von Daten aus den neu konzipierten Tests möglich sind.

Timing des Testentwicklungsprozesses

Abschnitt 1.3.3 stellt dar, wie die Aufgabenentwicklung vor allem im Hinblick auf die ÜGK 2024 wesentlich gestärkt wurde. Das Gutachten schlägt jedoch weitere Verbesserungen vor. Demnach soll nach der Präpilotierung und der Piloterhebung und vor der Schwellenwertsetzung zusätzlich eine Normierungsstudie durchgeführt werden, bei der in repräsentativen Stichproben nur fertig entwickelte Aufgaben eingesetzt werden. Nach Vorliegen des fertigen Aufgabensatzes und der statistischen Daten zu den Aufgaben und den Schülerinnen und Schülern soll die Schwellenwertsetzung stattfinden. Damit soll die Gefahr gebannt werden, dass aufgrund von Anpassungen an den Aufgaben zwischen Piloterhebung und Haupterhebung

die Schwellenwertsetzung erst nach der Haupterhebung stattfinden kann. Zu diesem Zeitpunkt besteht aus Sicht der Gutachtengruppe eine erhöhte Gefahr der politisch motivierten Einflussnahme auf Schwellenwerte.

In der Planung für die ÜGK 2024 ist die Piloterhebung als Normierungsstudie vorgesehen. Dies erschien nach einem wesentlichen Ausbau der Präpilotierung möglich. Allerdings mussten nach den Präpilotierungen relativ viele Aufgaben noch angepasst werden, ohne dass sie anschliessend noch einmal in einem Testkontext präpilotiert werden konnten. Damit besteht tatsächlich die Gefahr, dass aufgrund der statistischen Ergebnisse nach der Piloterhebung weitere Anpassungen nötig sind.

Der Kosta HarmoS teilt das Anliegen, dass Schwellenwertsetzungen nicht erst nach der Haupterhebung durchgeführt werden sollen. Die Notwendigkeit von Normierungsstudien soll geprüft werden. Bei der Umsetzung eines langfristig geplanten Designs, das Trendmessungen vorsieht, wird für die Schwellenwertsetzung ein neues Konzept entwickelt werden müssen. Die Vergleichbarkeit der Schwellenwerte von aufeinanderfolgenden Trendmessungen wird in erster Linie dadurch gewährleistet, dass Aufgaben aus demselben Aufgabenpool verwendet bzw. wiederverwendet werden.

Aufgabenpools

Die GS ADB hat sich bisher jeweils punktuell mit der Entwicklung der Tests für die ÜGK-Zyklen 2016, 2017, 2023 und 2024 befasst. Die Aufgaben der vergangenen Zyklen sind teils freigegeben und werden teils als Anker-Aufgaben für die Wiederverwendung in zukünftigen Zyklen unter Verschluss gehalten.

Das Gutachten weist darauf hin, dass bei der Entscheidung für ein Trenddesign für genügend solche Anker-Aufgaben gesorgt werden muss, damit Trendmessungen längerfristig sichergestellt sind.

Der Kosta HarmoS sieht die Notwendigkeit, für Trendmessungen genügend wiederverwendbare Aufgaben bereit zu halten und zu pflegen. Bei der Planung eines Aufgabenpools ist u.a. zu prüfen, ob die Pools für das 8. und das 11. Schuljahr durch gemeinsame Aufgaben miteinander verlinkt werden können, um bei den geplanten Kohortenlängsschnitten echte Fortschrittsmessungen zu ermöglichen. Dies wäre bei einem Testfokus allein auf den Grundkompetenzen kaum möglich. Wenn aber zu beiden Zeitpunkten Aufgaben für sehr schwache und sehr starke Schülerinnen und Schüler eingesetzt werden, könnten die Überlappungen genügen, um eine vertikale Verankerung der Skalen für die beiden Stufen zu rechtfertigen.

Schwellenwertsetzung

Wie im ersten Hauptteil beschrieben, wurden bei den bisherigen ÜGKs mittels einer modifizierten Bookmark-Methode die Schwellenwerte zwischen «Grundkompetenzen erreicht» und «Grundkompetenzen nicht erreicht» ermittelt.

Das gewählte Vorgehen entspricht gemäss der Gutachtengruppe dem *state of the art*.

Wenn in Zukunft die Ergebnisse nicht nur in zwei Kategorien aufgeteilt werden sollen, sondern in mehrere (je nach der Anzahl der Kompetenzstufen, die unterschieden werden sollen), dann muss die Schwellenwertsetzung entsprechend erweitert werden. Die Schwellenwerte, die in Zukunft für den Aufgabenpool gelten sollen, aus dem immer wieder Ankeraufgaben genommen werden, müssen sehr sorgfältig gesetzt werden.

Planung und Umsetzung der Kontextualisierung

Wie im ersten Hauptteil beschrieben, wurden in den bisherigen ÜGK-Erhebungen Fragebogen für Schülerinnen und Schüler eingesetzt, um hauptsächlich Zusammenhänge zwischen den Leistungen und erklärenden Variablen wie dem sozioökonomischen Hintergrund aufzeigen zu können. Fachdidaktische Interessen sowie spezifische Forschungsinteressen kamen hinzu. Eine qualitativ hochstehende

Kontextualisierung ist eine wichtige Voraussetzung für die Gewinnung von relevantem und verlässlichem Wissen für die Steuerung des Bildungssystems.

Das Gutachten macht an verschiedenen Stellen Aussagen dazu, wie die Kontextualisierung als vollwertiges Element der ÜGK konzipiert und umgesetzt werden sollte:

- Die Ziele und Funktionen der ÜGK bzw. eines ÜGK-Zyklus bilden den Ausgangspunkt für die Entwicklung von Kontextfragebögen, weil sich daraus die grossen Fragestellungen für eine Erhebung ergeben. Diese Fragestellungen können durch Fragestellungen für vertiefende Forschung ergänzt werden.
- Aus den Fragestellungen ergeben sich Anforderungen an das Stichprobendesign der ÜGK (z.B. Einbezug ganzer Klassen), das Erhebungsdesign und die Erhebungsinstrumente (z.B. Fragebogen für Schülerinnen und Schüler, bestimmte Fachlehrpersonen sowie Schulleitungen). Eventuell werden auch Verknüpfungen (über die AHVN13) mit anderweitig erhobenen Daten vorgesehen.
- Wenn bezüglich Zielen, Funktionen, Fragestellungen, Erhebungsdesign, Erhebungsinstrumenten und evtl. Datenverknüpfungen Konsens besteht, halten die bildungspolitischen Akteure das gemeinsam Vereinbarte im Sinne eines strategischen Entscheids fest.
- Als Grundlage für die Umsetzung der Kontextualisierung wird ein detalliertes Kontextualisierungs-Framework (ähnlich dem weiter oben thematisierten Test-Framework) ausgearbeitet, auf das sich Auftrags- und Umsetzungsebene einigen. Es beschreibt die zu erhebenden Konzepte sowie die Methoden und Instrumente. Bei den Instrumenten handelt es sich in der Regel um Fragebögen für eine oder mehrere Zielgruppen.
- Die weiteren Schritte setzen die wissenschaftlichen Auftragnehmenden gemäss wissenschaftlichen Gütekriterien um. Die bildungspolitischen Akteure werden mit den Ergebnissen bedient (zum Verhältnis von Auftragnehmenden und Auftraggebenden siehe 2.4.1).

Dem Kosta HarmoS erscheint es zielführend, wenn den ÜGK-Erhebungen ein gegenseitig verbindliches Gesamt-Framework zugrunde gelegt wird, das die Ziele und Mittel einer Erhebung insgesamt festhält.

Datennutzung für Sekundäranalysen

Die Abschnitte 1.3.4 und 1.3.5 geben einen kurzen Überblick zur Bereitstellung und Nutzung der ÜGK-Daten für Analysen, die über die Erstberichterstattung hinausgehen. Die Zahl solcher Analysen ist insgesamt überschaubar.

Die Gutachtengruppe äussert sich zur Beantragung und Nutzung der ÜGK-Daten. Ausgehend von der Feststellung, dass die ÜGK-Daten eher selten von Wissenschaftlerinnen und Wissenschaftlern, die nicht bereits Beziehungen zur ÜGK haben, beantragt werden, schlägt sie vor, die Daten in der Wissenschafts-community stärker zu bewerben sowie die Antrags- und Genehmigungsprozesse zu evaluieren und zu vereinfachen. Zudem sollen im Datennutzungskonzept noch vorhandene Unklarheiten ausgeräumt und für Konfliktfälle zwischen den Interessen der Wissenschaft und den genehmigenden Behörden ein Clearing-Verfahren geschaffen werden.

Aus den Antworten der schriftlichen Befragung, die Grundlage des Gutachtens war, geht der Wunsch hervor, dass die Dokumentation der ÜGK-Daten und der technischen Vorgehensweise verbessert und vereinfacht wird, damit sie auch für Forschende von ausserhalb der ÜGK besser zugänglich ist.

Wie auch aus Abschnitt 2.2 zum Erhebungsdesign hervorgeht, teilt der Kosta HarmoS das Anliegen der Gutachtengruppe, der Forschung attraktive Daten zur Verfügung zu stellen. Die Erfahrungen zeigen, dass sich an einer Datennutzung interessierte Forschende von ausserhalb des ÜGK-Projekts tatsächlich oft beraten lassen müssen, also nicht direkt aufgrund der Dokumentation zu den Scientific-Use-Files arbeiten können. Die Scientific-Use-Files sind zwar umfassend dokumentiert, aber Interessierte ohne genaue Projektkenntnisse dürften Mühe haben, sich selbständig zu orientieren. Interessierte sollten stärker «bei der Hand genommen» werden. Datenbestände aus den ÜGK-Erhebungen, die nicht in den Scientific-Use-

Files enthalten sind (z.B. die Originalantworten zu den Test- und Fragebogen-Items oder Prozessdaten, die von den digitalen Testsystemen aufgezeichnet wurden), dürften ausserhalb der ÜGK kaum bekannt sein. Die ÜGK-Daten sollten insgesamt bekannter und leichter zugänglich gemacht werden. Die Erfahrungen mit den Datennutzungskonzepten vom 22. März 2018 (EDK, 2018a) und vom 5. Mai 2021 (EDK, 2021b) sind bisher positiv; zu Streitfällen ist es im Zusammenhang damit nicht gekommen.

Dokumentation der ÜGK-Zyklen

Bezüglich der Dokumentation der ÜGK-Zyklen (vgl. 1.3.4) kritisiert das Gutachten eine zu grosse Heterogenität. Mehr Einheitlichkeit in den technischen Berichten und Dokumentationen über die Sprachregionen, Testdomänen und Entwicklungsetappen hinweg würde die Rezeption und die Beurteilung erleichtern. Das Gutachten schlägt vor, pro Testzeitpunkt nur noch einen einzigen technischen Bericht zu erstellen und zudem künftige Berichte nach einer einheitlichen Vorlage (Template) zu verfassen. Dabei ist es u.a. wichtig, die Daten und die technische Vorgehensweise für hochwertige Forschungsarbeiten präzise genug zu beschreiben.

Nach Einschätzungen des Kosta HarmoS sind diese Empfehlungen zielführend.

Berichterstattung und Kommunikation der ÜGK

Aus den Befragungen zum Gutachten geht hervor, dass die meisten Akteurinnen und Akteure die Berichte zur ÜGK als qualitativ gut einschätzen. Verbesserungsbedarf wird dagegen bei der Kommunikation der Ergebnisse gesehen. Das Gutachten empfiehlt, die Kantone rechtzeitig und aktiver in die Kommunikation einzubeziehen, nicht zuletzt um die Sichtbarkeit der ÜGK in der Öffentlichkeit zu stärken. Die Kantone sollten der nationalen Berichterstattung nicht hinterherhinken müssen.

Von Kantonsseite wird auch mehr fachliche Unterstützung aus dem ÜGK-Projekt (ICER, Durchführungszentren) gewünscht, einerseits bei der Interpretation der Ergebnisse, andererseits in Bezug auf Erläuterungen zum Kontextfragebogen. Gute Erklärungen können helfen, der teilweise mangelnden Akzeptanz von bestimmten Fragen in der Politik und in der Bevölkerung zu begegnen.

Das GS EDK hat bereits in Zusammenhang mit der Erstberichterstattung zu den Ergebnissen der ÜGK 2016 und 2017 im Mai 2019 aktiv kommuniziert und hat den Kantonen Informationsangebote gemacht, die diese aber sehr unterschiedlich in Anspruch nahmen. Dem Kosta HarmoS ist es grundsätzlich ein grosses Anliegen, dass die Erstkommunikation zu den ÜGK-Erhebungen so vorbereitet wird, dass sie die verdiente Aufmerksamkeit erhält. Die Kantone sollen dabei konsequent als privilegierte Partner behandelt werden. Zu diesem Zweck werden die Kantone proaktiv mit Informationen und Informationsangeboten (auch fachlich-inhaltlichen der PL ÜGK) versorgt. Als Kommunikationskanäle werden besonders die verschiedenen Gremien und Gruppen der EDK genutzt, zudem die Kontakte der Kommunikationsabteilung der EDK.

Aus den Befragungen zum Gutachten geht weiter hervor, dass bei der Kommunikation, die sich an die Lehrpersonen richtet, Verbesserungsbedarf besteht. Insbesondere sollten für die Lehrpersonen und die Schülerinnen und Schüler massgeschneiderte Informationsmaterialien zur Verfügung gestellt werden, um die ÜGK bekannt zu machen und für die ÜGK Verständnis zu wecken. Die Materialien sollten Folgendes leisten: über die Tests informieren (u.a. mit Aufgabenbeispielen); die Erwartungen an Lehrpersonen und Schülerinnen und Schüler klären; erklären, warum es keine individuellen Rückmeldungen gibt.

Eine gute Kommunikation mit den Schulleitungen, Lehrpersonen und den Schülerinnen und Schülern erachtet der Kosta HarmoS gerade im Hinblick auf die künftig grössere Zahl an ÜGK-Erhebungen als zunehmend wichtig. Es ist zu erwarten, dass durch eine gute, anschauliche Kommunikation letztlich auch die Testmotivation gesteigert werden kann. Mangelhafte Motivation und Kooperationsbereitschaft bei den Erhebungen kann zu einer Verzerrung der Ergebnisse führen.

Gemäss den Befragungen fühlen sich die kantonalen Referenzpersonen für Large-Scale-Assessments insgesamt gut über die ÜGK informiert, insbesondere auch dank den sprachregionalen Durchführungszentren. Die Sitzungen der kantonalen Referenzpersonen werden nicht immer als nützlich empfunden: Die Informationen sind zum Teil schon bekannt; die Gelegenheit zu diskutieren wird oft nicht wahrgenommen; die Informationen der PL ÜGK kommen teilweise zu wissenschaftlich daher.

Die im Gutachten genannten Bedürfnisse der Mitglieder der Arbeitsgruppe der kantonalen Referenzpersonen für Large-Scale-Assessments sind dem GS EDK bekannt, und es wird versucht, die Sitzungsgestaltung dementsprechend anzupassen.

Die Gutachtengruppe selbst weist darauf hin, dass in der Erstberichterstattung nicht mit statistisch adjustierten Werten zum Erreichen der Grundkompetenzen gearbeitet sollte, sondern nur mit den beobachteten. Die Erstberichterstattung berichtet jeweils pro Kanton bzw. Kantonsteil (bei zwei- oder mehrsprachigen Kantonen) den Anteil der Schülerinnen und Schüler, welche die Grundkompetenzen erreicht haben. Zudem zeigt sie sog. adjustierte Vergleiche, die der unterschiedlichen sozialen Zusammensetzung der Schülerschaft in den Kantonen Rechnung tragen. Die Gutachtengruppe befürchtet, dass solche Anpassungen den Handlungsbedarf relativieren, indem sie ihn «wegerklären». Dabei sollten Massnahmen gerade da ausgelöst werden, wo soziale Faktoren den schulischen Erfolg beeinträchtigen.

Für den Kosta HarmoS wäre dieser Effekt des «Wegerklärens» selbstverständlich unerwünscht. Die Erstberichterstattung muss im Hinblick auf die Neukonzipierung der ÜGK ohnehin angepasst werden. Wenn das Zurverfügungstellen von Erklärungsansätzen für Unterschiede zwischen den Ergebnissen neben der Systemevaluation zu einem vollwertigen Ziel der ÜGK werden soll (vgl. 2.1.2), wird sich das auch in der Berichterstattung niederschlagen. Soziodemografische Faktoren werden dabei ein Aspekt unter anderen sein.

Schnittstelle zum Bildungsbericht

Ergebnisse der ÜGK können grundsätzlich auf drei Wegen in den Bildungsbericht einfliessen: 1) direkt (insbesondere die Kantonsvergleiche, die es sonst nicht gibt); 2) durch Vertiefungsstudien der SKBF auf Grundlage der ÜGK-Daten zu Thematiken des Bildungsberichts; 3) über Forschungsarbeiten anderer, die auf der ÜGK aufbauen.

Die Ergebnisse der Erstberichterstattung der bisherigen ÜGK-Zyklen konnten und können wegen Problemen bei der zeitlichen Abstimmung nicht in den Bildungsbericht aufgenommen werden. Vertiefende Untersuchungen zu ÜGK-Daten sind aber in den Bildungsbericht 2023 eingearbeitet.

Das vorgeschlagene längerfristig fest geplante Kohortenlängsschnitt-Design mit regelmässigen Trendmessungen über das ganze Leistungsspektrum hinweg kommt laut dem Direktor der SKBF den Bedürfnissen der Bildungsberichterstattung stark entgegen:

- Die Nutzung von ÜGK-Daten wird planbar. ÜGK-Daten sind wertvoll, weil sie (fast) alle Kantone mit repräsentativen Stichproben oder Vollerhebungen umfassen.
- Die Testskalen mit zuverlässigen Messungen über das ganze Kompetenzspektrum eignen sich, um statistische Zusammenhänge zwischen Testdaten und Kontextdaten zu untersuchen.
- Trendmessungen sind dafür konzipiert, Steuerungswissen zu generieren.
- Die Kohortenlängsschnitte umfassen in kleineren Kantonen auch zahlreiche individuelle Längsschnitte. Diese eignen sich besonders, um Entwicklungsverläufe nachzuzeichnen.

Um Einflussfaktoren im Kontext der Schule korrekt zu modellieren, wäre es zudem hilfreich, wenn Klassenstichproben gezogen und Kontextmerkmale auf Ebene der Klasse und des Unterrichts erhoben würden.

Anhang 2: Gutachten zur ÜGK

Universität Zürich[UZH]

Institut für Erziehungswissenschaft

Wissenschaftliches Gutachten zur „Governance der Überprüfung des Erreichens der Grundkompetenzen (ÜGK)“

Prof. Dr. Katharina Maag Merki (Universität Zürich), Leitung

Dr. Anja Giudici (Universität Oxford)

Prof. i.R. Dr. Eckhard Klieme (DIPF – Leibniz-Institut für Bildungsforschung und Bildungsinformation, Johann Wolfgang Goethe-Universität Frankfurt am Main.)

Prof. Dr. Hans-Anand Pant (Humboldt-Universität zu Berlin)

Prof. Dr. Claudia Schreiner (Universität Innsbruck)

Mitarbeit und Koordination: Kai Schudel (Universität Zürich)

31. Oktober 2021

Inhaltsverzeichnis

1 Einleitung **4**
1.1 Strukturierung der Fragen und Ergebnisdarstellung 4
1.2 Informationsgrundlage 5

2 Aufgaben und Ertrag der ÜGK **8**
2.1 Ziele und Funktionen der ÜGK (A1.2) 9
2.2 Nutzung der Ergebnisse für den eigenen Arbeits- und Zuständigkeitsbereich (A1.3) 11
2.3 Umfang der Evidenz im Gesamtsystem (A2.8; A2.9) 14
2.4 Kommunikation der Ergebnisse der ÜGK und Berichterstattung (A1.4; A1.5) 15
2.5 ÜGK und Bildungsbericht (A2.10) 16

3 Zielvorgaben der ÜGK **18**
3.1 Auswahl der Fächer und Jahrgangsstufen (B1.1; B1.2) 18
3.2 Mindeststandards oder differenzierte Leistungsstufen? (B1.4; B3.13; B3.15) 20

4 Design der ÜGK **29**
4.1 Erhebungsrhythmus und Stichprobenziehung (B1.3; B1.5; B1.6; B2.8) 29
4.2 Konzeption und Entwicklung der Kontextfragebögen (B3.16) 32
4.3 Prozesse der Aufgaben- und Testentwicklung (B3.11; B3.12) 34
4.4 Technische Fragen zur ÜGK (B3.14; B3.17; B3.18) 37

5 Institutionelle Gestaltung der ÜGK **39**
5.1 Organisation und Verantwortungsstruktur (B2.7; B.2.8; B2.9) 39
5.2 Verhältnis ÜGK und Wissenschaft (A1.6; B2.10; B3.16; B3.19) 43

6 Empfehlungen **47**
6.1 ÜGK im Gesamtsystem 49
6.1.1 Weiterführung der ÜGK bei substanziellen Anpassungen im Gesamtdesign, der institutionellen Verankerung und Gestaltung sowie in der Zuweisung von finanziellen und personellen Ressourcen 49
6.1.2 Langfristige Planung bezüglich Erhebungsdesign gewährleisten und Durchführung vertraglich verbindlich absichern 49
6.1.3 Verhältnis der ÜGK zu Monitoringinstrumenten in den Kantonen und Sprachregionen klären 50
6.1.4 Aufgaben und Rollen von Politik, Bildungsadministration und Forschung/Wissenschaft im Gesamtgefüge der ÜGK klären 50

6.2 Grundlagen der ÜGK 52
6.2.1 Funktionen der ÜGK klären 52
6.2.2 Frage nach den Grundkompetenzen klären 52
6.2.3 Frage nach der Harmonisierung der Grundkompetenzen klären 53
6.3 Zielvorgaben der ÜGK 54
6.3.1 An einer Orientierung an Mindeststandards und der Überprüfung des Erreichens der Grundkompetenzen festhalten sowie die 2011 formulierten Grundkompetenzen weiterhin als Basis der Aufgabenentwicklung akzeptieren 54
6.3.2 Frage klären, welche Bedeutung der soziale Vergleich zwischen den Kantonen hat und adäquate Lösungen für den Kantonsvergleich finden 55
6.4 Design der ÜGK 56
6.4.1 An der Wahl der Jahrgangsstufen und der Fächer / Fachbereiche sowie an einer standardisierten Erhebung über alle Kantone hinweg festhalten 56
6.4.2 Erhebungsdesign klären: Trend-, Kohorten- und Längsschnitterhebungen? 56
6.4.3 Inhalte der nationalen Berichte klären 57
6.4.4 Kontextinformationen erweitern, in Abhängigkeit der Funktionen der ÜGK 57
6.4.5 Stichprobenziehung entsprechend den Funktionen der ÜGK anpassen 58
6.4.6 Öffentlichen Kommunikation der Ergebnisse der ÜGK verbessern 58
6.4.7 Weitere Empfehlungen 58
6.5 Institutionelle Gestaltung der ÜGK 59
6.5.1 Klare sowie vereinfachte Zuständigkeits-, Entscheidungs- und Verantwortungsstrukturen implementieren 59
6.5.2 Testentwicklung als komplexes und anspruchsvolles wissenschaftliches Aufgabenfeld anerkennen, mit genügend personellen Ressourcen ausstatten und institutionell in die wissenschaftliche Projektstruktur der ÜGK einbinden 60
6.5.3 Fachdidaktik(en) stärker in den Gesamtprozess der ÜGK einbinden 61
6.5.4 Sprachregionale Unterschiede in Bezug auf fachliches Lernen und Aufgaben herausarbeiten sowie für die Testentwicklung nutzen 61

I. Anhang – Literaturverzeichnis 62

II. Anhang – Ergebnisse der standardisierten Befragung 64

III. Anhang – Fragekatalog der EDK 82

1 Einleitung

Im Januar 2021 hat die EDK beschlossen, wissenschaftliche Expert*innen mit einem Gutachten zu fachlichen Aspekten der „Governance der Überprüfung des Erreichens der Grundkompetenzen (ÜGK)“ zu beauftragen. Der Auftrag wurde einer internationalen Gutachtengruppe, bestehend aus Dr. Anja Giudici (Universität Oxford), Prof. i.R. Dr. Eckhard Klieme (DIPF), Prof. Dr. Hans Anand Pant (Humboldt-Universität zu Berlin) und Prof. Dr. Claudia Schreiner (Universität Innsbruck) unter der Leitung von Prof. Dr. Katharina Maag Merki (Universität Zürich) erteilt. Die Gutachtengruppe hat das vorliegende Gutachten im Zeitraum von April bis Oktober 2021 auf Basis eines Fragekatalogs der EDK erarbeitet.

1.1 Strukturierung der Fragen und Ergebnisdarstellung

Die Ausgangslage des wissenschaftlichen Gutachtens bildet der Fragekatalog der EDK (siehe Anhang III), der grob in einen Bereich A der „externen Governance“ und einen Bereich B der „internen Governance“ aufgeteilt ist. Der Bereich A wird in die zwei Teilbereiche „A1 Systemsteuerung, Zielkohärenz und Handlungskoordination“ und „A2 Die ÜGK und weitere Erhebungen im Bildungssystem“ unterteilt. Der Bereich B umfasst drei Teilbereiche: „B1 Erhebungsdesign“, „B2 Projektorganisation und institutionelle Verankerung“ sowie „B3 Fachliche und technische Aspekte“.
Eine Analyse der Fragestellungen hat gezeigt, dass einzelne Aspekte, so zum Beispiel der Kontextfragebogen oder das Verhältnis von Wissenschaft und Politik, an verschiedenen Stellen angeschnitten werden. Deshalb hat die Gutachtengruppe entschieden, die Fragen thematisch zu sortieren und in einem groben Modell darzustellen (Abbildung 1). In einem ersten Bereich sind Fragen zu den *Aufgaben der ÜGK im Bildungsmonitoring* angesiedelt. Die Definition der Aufgaben sowie die Positionierung der ÜGK im Gesamtsystem führt in der Folge zu *Zielvorgaben der ÜGK* bezüglich Fächerauswahl und Definition von Bildungsstandards. Diese Zielvorgaben wiederum münden in eine operative Umsetzungs-Ebene mit Implikationen für das *Design der ÜGK* und für *die institutionelle Gestaltung der ÜGK*. Der *Ertrag der ÜGK* hängt wiederum zusammen mit den allgemeinen Vorstellungen über die Aufgaben der ÜGK im Bildungsmonitoring und die damit verbundenen Erwartungen an die ÜGK. Darüber hinaus ist der wahrgenommene Ertrag abhängig von der konkreten Umsetzung der ÜGK.

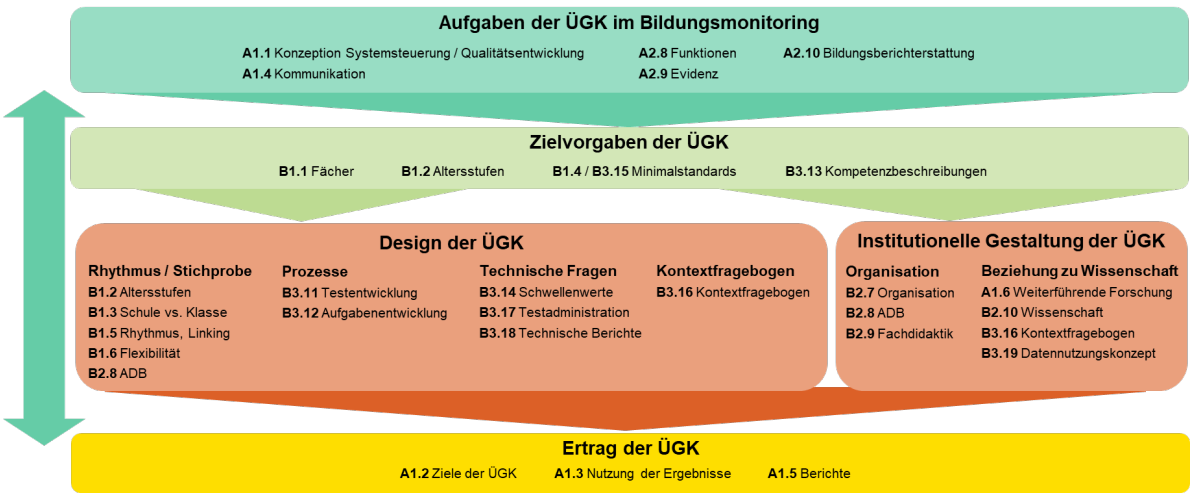


Abbildung 1: Modell zur Analyse der ÜGK im Gutachten

Die Darstellung der Befunde der Gutachtengruppe orientiert sich am obigen Modell (Abbildung 1). In Kapitel 2 werden zuerst Befunde zu *Aufgaben und Ertrag der ÜGK* zusammengefasst, Kapitel 3 präsentiert die Befunde zu den *Zielvorgaben der ÜGK*, Kapitel 4 berichtet die Befunde zum *Design der ÜGK* und in Kapitel 5 werden die Befunde zur *institutionellen Gestaltung der ÜGK* erläutert. Diese Kapitel werden

jeweils in Themen unterteilt, die in den Untertiteln beschrieben sind und gleichzeitig auf die Nummern/spezifischen Fragen des Fragekatalogs der EDK verweisen. Die Fragen A1.1 und A1.7 zielen auf Empfehlungen ab und werden entsprechend nicht explizit unter Befunde erörtert. [1] In Kapitel 6 schliesslich sind Empfehlungen der Gutachtengruppe zusammengefasst, die in die Bereiche *ÜGK im Gesamtsystem (Kapitel 6.1)*, *Grundlagen der ÜGK (Kapitel 6.2)*, *Zielvorgaben der ÜGK (Kapitel 6.3)*, *Design der ÜGK (Kapitel 6.4)* und *institutionelle Gestaltung der ÜGK (Kapitel 6.5)* aufgeteilt sind.

1.2 Informationsgrundlage

Die Ergebnisse und Empfehlungen des Gutachtens basieren auf Dokumenten der ÜGK, auf Einzel- und Gruppengesprächen mit Akteur*innen, einer standardisierten Befragung von Akteur*innen und auf der Fach-Expertise der Gutachter*innen. Das Generalsekretariat der Schweizerischen Konferenz der Erziehungsdirektoren (GS EDK) hat den Gutachter*innen für den Bearbeitungszeitraum Zugang zu allen relevanten Dokumenten zur ÜGK gewährt (z.B. Verträge, Berichte, Konzeptpapiere). Auf dieser Basis konnte eine *Dokumentenanalyse* vorgenommen werden.
Auf Grundlage der Dokumentenanalyse, sondierender Gespräche mit Akteur*innen der ÜGK und dem Fragekatalog der EDK wurden in einem nächsten Schritt für das Gutachten *relevante Akteur*innen definiert*. Orientiert an der im Fragekatalog angelegten Unterscheidung zwischen externer und interner Governance wurde zwischen externen und internen Akteur*innen unterschieden.
*Externe Akteur*innen* sind aktuelle oder mögliche Adressat*innen der ÜGK-Ergebnisse, die also Ansprüche an die Nutzung der ÜGK artikulieren können. Als wesentliche externe Akteur*innen wurden die Vertretungen der kantonalen Bildungsadministrationen und der EDK-Regionalkonferenzen sowie die Bildungsforschung und die Schulpraxis identifiziert. Die Bildungsadministrationen können als zentrale Stakeholder betrachtet werden, da das Bildungsmonitoring sowie die Qualitätsentwicklung im Bildungssystem in ihrem Verantwortungsbereich liegen. Hier wurden einerseits kantonale Vertreter*innen fokussiert, die Auskunft darüber geben können, ob und wie ÜGK-Ergebnisse für das kantonale Bildungsmonitoring und die Qualitätsentwicklung nutzbar sind, andererseits auch Personen, die mit der Umsetzung der ÜGK beauftragt sind (kantonale Referenzpersonen). Es hat sich gezeigt, dass diese beiden Funktionen oft personell dort zusammenfallen, wo kantonale Bildungsdirektionen (oder Volksschulämter) in Bezug auf Grösse und Ressourcen gering ausgestattet sind. Entsprechend wird die Perspektive dieser beiden Gruppen im Folgenden nicht mehr differenziert und als Perspektive der kantonalen Vertreter*innen (oder der „kantonalen Vertretungen“) zusammengefasst. Als *interne Akteur*innen* wurden Personen adressiert, die an der Gestaltung und Umsetzung der ÜGK beteiligt sind. Dies sind die Verantwortlichen der Projektleitung ÜGK, das Board und der Beirat des Interfaculty Centre for Educational Research (ICER) der Universität Bern, Mitglieder des Kosta HarmoS und externe Wissenschafter*innen, die im Rahmen des wissenschaftlichen Konsortiums an der ÜGK mitarbeiten. Teilweise sind diese Akteur*innen in mehreren Rollen in die ÜGK involviert.
Die Erfahrungen und Sichtweisen dieser Akteur*innen wurden für das Gutachten über zwei Wege eingeholt. Zum einen wurden Einzel- und Gruppengespräche geführt, dies im Zeitraum Juni bis Oktober 2021, zum anderen wurden dieselben Personen gebeten, an einer standardisierten Befragung teilzunehmen. Diese Befragung fand im Juli 2021 statt. Die *Einzel- und Gruppengespräche* wurden meistens von jeweils zwei Gutachter*innen geführt, protokolliert und aufgenommen, damit die gesamte Gutachtengruppe die Gesprächsinhalte nachträglich nachvollziehen konnte[2]. Dabei wurden die Gruppen so zusammengestellt, dass die Akteur*innen ähnliche Funktionen repräsentieren und sprachregional zusammenpassten. Es konnten 12 Einzel- oder Gruppengespräche mit externen Akteur*innen geführt werden: mit Vertreter*innen von Kantonen aus den vier Landessprachen sowie mit Vertreter*innen der verschie-

[1] Die Frage A2.11 nach dem optimalen Beitrag der ÜGK zur Dateninfrastruktur konnte im Gutachten nicht untersucht werden.

[2] Sämtliche Tonaufnahmen werden nach Ende des Auftrags an die Gutachtengruppe gelöscht.

denen Regionalkonferenzen beziehungsweise von Kosta HarmoS. Dabei ist es – mit Ausnahme der Kantone Zug, Schaffhausen und der französischsprachigen Vertretung des Kantons Bern – gelungen, die Perspektive sämtlicher Kantone und Sprachregionen mindestens über eine Vertretung einzuholen. Wichtig ist, dass auch kleinere Kantone und solche, die dem HarmoS-Konkordat nicht beigetreten sind, vertreten waren. Zusätzlich nahmen in sechs Einzel- oder Gruppengesprächen auch externe Akteur*innen der Bildungsforschung und der Schulpraxis unterschiedlicher Sprachregionen teil. Interne Akteur*innen, die Projektleitung ÜGK und das wissenschaftliche Konsortium, wurden in sechs Einzel- oder Gruppengesprächen befragt und mit Fachdidaktiker*innen konnten zusätzlich drei Gespräche geführt werden. Ebenfalls wurden zwei Gespräche mit Mitgliedern des Beirats und des Boards geführt. Die Zusammenstellung aller Akteur*innen, deren Perspektiven in die Gespräche einflossen (Tabelle 1), zeigt, dass die Bereitschaft der Akteur*innen, das Gutachten mit ihrer Erfahrung zu unterstützen, sehr gross war, und dass es den Gutachter*innen gelungen ist, eine grosse Breite und Vielfalt an wichtigen Akteur*innengruppen zu befragen.

Externe Akteur*innen	Anzahl		Institutionen
	fr/it	dt/ro	
Vertreter*innen der Kantone, Regionalkonferenzen oder Kosta HarmoS	8	27	Alle Kantone (ausser ZG, SH und französischsprachiges BE) und alle Regionalkonferenzen
Schulpraxis	2	2	VSLCH, LCH, SER, CLACESO
Bildungsforschung		1	SKBF
Interne Akteur*innen	**Anzahl**		**Organisationen**
Akteur*innen der Projektleitung ÜGK	9		ICER, ADB, SUPSI, SRED, PHSG, IBE
Externe Wissenschafter*innen des Wissenschaftlichen Konsortiums	1		Arbeitsgruppe Kontextfragebogen
Mitglied des Beirats	1		
Mitglieder des Boards	3		
Fachdidaktiker*innen	4		

Tabelle 1: Personenanzahl der Akteur*innen für Einzel- und Gruppengespräche; SKBF=Schweizerischen Koordinationsstelle für Bildungsforschung; VSLCH=Verband Schulleiterinnen und Schulleiter Schweiz; LCH=Dachverband Lehrerinnen und Lehrer Schweiz; SER=Syndicat des Enseignantes et Enseignants de Suisse romande; CLACESO=Conférence Latine des Chefs d'Etablissement de la Scolarité Obligatoire

Der *standardisierte Fragebogen* wurde per E-Mail an dieselben Akteur*innen geschickt, die schon für die Einzel- und Gruppengespräche angeschrieben worden waren. Zusätzlich diente die Befragung dazu, die Einschätzungen von Forschenden aus der Bildungsforschung (z.B. Erziehungswissenschaft, Fachdidaktik) und von Personen aus der Aus- und Weiterbildung von Lehrpersonen in standardisierter Form zu erhalten. Für den Zugang zu dieser Akteur*innengruppe konnte auf den E-Mail-Verteiler der Schweizerischen Gesellschaft für Bildungsforschung (SGBF) zurückgegriffen werden. I
nsgesamt haben 141 Personen den Online-Fragebogen ausgefüllt. Es haben 39 kantonale Vertreter*innen teilgenommen, wovon 21 aus Kantonen stammen, die im HarmoS-Konkordat sind, und 15 aus Kantonen, die dem HarmoS-Konkordat nicht beigetreten sind. Damit kann die Datenbasis für die Kantone als belastbar interpretiert werden. Weitere 77 Personen, die den Fragebogen ausgefüllt haben, gaben an, in der Forschung und/oder der Aus- und Weiterbildung von Lehrpersonen tätig zu sein[3]. Bezüglich

[3] Die restlichen 25 Personen haben andere Funktionen angegeben, darunter 3 Personen aus der Schulpraxis.

der sprachregionalen Verteilung konnten 103 deutsch- oder romanischsprachige, 25 französischsprachige und 13 italienischsprachige Akteur*innen gewonnen werden. Die Befunde der standardisierten Befragung werden teilweise im Bericht präsentiert und stehen im Anhang im ganzen Umfang zur Verfügung.

Dank

An dieser Stelle möchten wir unseren Dank allen Personen aussprechen, die die Erstellung dieses Gutachtens unterstützt haben.

Zunächst geht unser Dank an alle Personen, die ihre Erfahrungen und Einschätzungen mit uns geteilt und sich für ein Einzel- oder Gruppengespräch zur Verfügung gestellt haben. Ebenso gebührt unser Dank all jenen, die die standardisierte Befragung mitten in der Sommerpause bearbeitet haben. Sehr herzlich möchten wir zudem Peter Lenz vom Generalsekretariat der EDK danken, der uns mit allen notwendigen Unterlagen ausgestattet hat und uns in unserer Arbeit hervorragend unterstützt hat. Zu guter Letzt gebührt unser Dank Susanne Hardmeier, Generalsekretärin der EDK, sowie der EDK, die uns die Erstellung des Gutachtens anvertraut haben.

Wir würden uns sehr freuen, wenn das Gutachten einen Beitrag leistet, das Bildungsmonitoring in der Schweiz und die Realisierung der ÜGK qualitativ weiterzuentwickeln.

2 Aufgaben und Ertrag der ÜGK

Die ÜGK ist als Instrument der Systemevaluation im Rahmen des nationalen Bildungsmonitorings eingebettet in ein komplexes Set an weiteren Leistungsüberprüfungen und Standortbestimmungen, die mit unterschiedlicher Intensität an obligatorischen Schulen in der Schweiz durchgeführt werden.

- Die Schweiz nimmt seit 2000 an der internationalen PISA-Studie teil, seit 2015 – im Zusammenhang mit dem Entscheid zur ÜGK – ausschliesslich mit einer nationalen Stichprobe (mit Ausnahme des Kantons Tessin, der nach wie vor mit einer kantonal repräsentativen Stichprobe an den PISA-Erhebungen teilnimmt). Die Teilnahme an der PISA-Studie erlaubt die Positionierung der Schweiz im internationalen Vergleich, ermöglicht aber keine kantonsvergleichenden Analysen mehr.
- Mit der Einführung der ÜGK liegt nun erstmals ein nationales, auf schweizerische Bedürfnisse ausgerichtetes Instrument der Systemevaluation vor, welches diese kantonsvergleichenden Analysen mit explizitem Bezug zu den entwickelten Bildungsstandards möglich macht. Alle Kantone – mit Ausnahme des Kantons Zug ab 2023 – nehmen an diesen Erhebungen teil, auch jene, die nicht dem HarmoS-Konkordat beigetreten sind.
- Gemäss Umsetzungsstrategie des HarmoS-Konkordates (EDK, 2007, Art. 5.3.1 und 5.3.2) sollen, ebenfalls auf Basis der Bildungsstandards und der regionalen Lehrpläne, Instrumente zur individuellen Standortbestimmung der Schüler*innen auf sprachregionaler Ebene entwickelt werden. Dieses Ziel ist allerdings in der geplanten Form (noch) nicht realisiert worden.
- In einzelnen Fällen, beispielsweise in der Nordwestschweiz, kommen kantonsübergreifende Erhebungen und Leistungstests zum Einsatz, wobei der Bezug zu den Mindeststandards teils unklar ist.
- Zusätzlich haben rund die Hälfte der Kantone eigene standardisierte Leistungstests auf der Primarstufe und fast alle Kantone standardisierte Leistungstests auf der Sekundarstufe implementiert, wobei der Bezug zu den Mindeststandards teils unklar ist.[4]
- In verschiedenen Kantonen werden keine eigenen Leistungstests durchgeführt.

Damit variiert die Erhebungsdichte je nach Kanton stark. Zudem haben die in den Kantonen oder Regionen eingesetzten Instrumente unterschiedliche Ziele. Zwar dienen die meisten in den Kantonen und Regionen angewendeten Tests primär der Standortbestimmung der Schulen, Klassen oder Schüler*innen, wie beispielweise die in verschiedenen deutschsprachigen Kantonen angewendeten Stellwerk- oder Klassencockpit-Tests oder die Genfer EVACOM. Allerdings dienen verschiedene Leistungsevaluationen zusätzlich auch dem kantonalen oder regionalen Systemmonitoring, darunter beispielweise die Épreuves Cantonales de Référence (ECR) im Kanton Neuenburg oder die in der Nordwestschweiz zum Einsatz kommenden Checks.[5] Diese Situation trägt dazu bei, dass das Bedürfnis nach Daten und Analysen zur Systemevaluation auf nationaler Ebene je nach Kanton variiert. Die in den Gesprächen angesprochenen Problemlagen in Bezug auf Ziele und Funktionen der ÜGK in Abstimmung mit anderen Erhebungsinstrumenten müssen vor dem Hintergrund dieser heterogenen Erhebungs- und Datendichte gelesen werden.

[4] Siehe dazu die entsprechende EDK Kantonsumfrage von 2019/20: https://www.edk.ch/edk.ch/platform/de/de/bildungssystem/kantonale-schulorganisation/kantonsumfrage/b-35-standardisierte-leistungstests

[5] https://www.edk.ch/edk.ch/platform/de/de/bildungssystem/kantonale-schulorganisation/kantonsumfrage/b-35-standardisierte-leistungstests

2.1 Ziele und Funktionen der ÜGK (A1.2)

Breite Akzeptanz des Ziels „Überprüfung des Erreichens der Grundkompetenzen“ als Systemevaluation auf nationaler Ebene, unabhängig davon, ob der eigene Kanton dem HarmoS-Konkordat beigetreten ist oder nicht.

Die ÜGK ist heute das einzige Instrument, das eine interkantonal vergleichende Lernstandsanalyse erlaubt. Das entsprechende Ziel, die Überprüfung des Erreichens der Grundkompetenzen in der Schweiz im Rahmen der nationalen Systemevaluation, wird in den Gesprächen durchgehend in hohem Masse akzeptiert. Die standardisierte Befragung zeigt, dass seitens der kantonalen Vertreter*innen alle Personen, mit einer Ausnahme, es eher wichtig (44 %) oder sehr wichtig (51 %) finden, das Erreichen der Grundkompetenzen der Schüler*innen auf nationaler Ebene zu überprüfen, dies unabhängig davon, ob der eigene Kanton dem HarmoS-Konkordat beigetreten ist oder nicht. Bei Personen aus der Forschung und Aus- und Weiterbildung ist die Zustimmung mit 74 % („eher wichtig“ oder „sehr wichtig“) ebenfalls sehr hoch. Von den kantonalen Vertretungen wird die ÜGK vor allem als (eher) wichtig für das Fällen von bildungspolitischen Entscheiden auf nationaler Ebene (90 %) beurteilt; auch hier bestehen keine Unterschiede zwischen den HarmoS- und Nicht-HarmoS-Kantonen. Die ÜGK wird darüber hinaus als eher wichtig oder sehr wichtig für das Fällen von bildungspolitischen Entscheiden auf kantonaler (77 %) und sprachregionaler Ebene (64 %) angesehen. Von einzelnen Vertreter*innen der Schulpraxis wird allerdings darauf hingewiesen, dass die ÜGK tendenziell ein dysfunktionales Potenzial für Konkurrenzverhalten und Ranking habe und die Durchführung nationaler Leistungserhebungen somit grundsätzlich wenig sinnvoll sei.

Der ÜGK werden neben der Systemevaluation auf nationaler Ebene verschiedene weitere Funktionen zugeschrieben.

Die Umschreibung der Funktion der ÜGK im Rahmen der Systemevaluation auf nationaler Ebene in den Gesprächen kann drei Themenbereichen zugeordnet werden, wobei die ersten beiden typische Ziele des Bildungsmonitorings sind, die auch im Zusammenhang mit anderen Bildungssystemen genannt werden:

- Interkantonaler Vergleich, d.h. soziale Bezugsnormorientierung: „Wo steht der eigene Kanton im Vergleich zu den anderen?“
- Harmonisierung des Erreichens der Grundkompetenzen zwischen den Kantonen und Sprachregionen: „Inwiefern ist die Harmonisierung realisiert?“
- Kriteriale Bezugsnormorientierung: „Haben die Schüler*innen in der Schweiz und im eigenen Kanton die Grundkompetenzen erreicht?“

In einzelnen Fällen erhoffen sich Vertreter*innen der Kantone durch die ÜGK zusätzlich eine Überprüfung von Massnahmen, die im Kanton in den letzten Jahren eingeführt wurden. Dies auch im Zusammenhang mit den Ergebnissen der PISA-Studien. Zudem könne die ÜGK die Funktion haben, einen überkantonalen Meinungsaustausch zu Bildungsthemen zu ermöglichen. Die Publikation der ÜGK-Resultate hätte das Potential, als notwendig erachtete regionale oder gar schweizweite Diskussionen über die Ziele und Strukturen des Bildungswesens zu entfachen. Um diese zu erfüllen, fehle es der ÜGK allerdings an Visibilität.

Die Befunde der standardisierten Befragung zeigen darüber hinaus, dass neben der Systemevaluation auf nationaler und kantonaler Ebene auch erwartet wird, dass *mittels der ÜGK Ursachen für Leistungsunterschiede zwischen regionalen und kantonalen Bildungssystemen identifiziert werden können*. 80 % der befragten kantonalen Vertreter*innen sowie 68 % der Personen aus der Forschung und Aus- und Weiterbildung geben an, dass dies eine relevante Funktion der ÜGK sei.

Umstrittener ist, ob die ÜGK Handlungswissen bereitstellen kann und soll.

In der standardisierten Befragung ist die Mehrheit der kantonalen Vertreter*innen der Ansicht, dass die ÜGK auch Handlungswissen bereitstellen sollte, wie das Bildungssystem auf nationaler (85 %), sprachregionaler (62 %) oder kantonaler (74 %) Ebene verändert werden kann. Etwas weniger häufig, aber mit 54% immer noch mehr als die Hälfte der kantonalen Vertreter*innen, stimmen der Aussage zu, dass die ÜGK auch die Funktion haben sollte, Handlungswissen darüber zu generieren, wie die Schulpraxis verändert werden kann. In den Gesprächen wird allerdings deutlich, dass für die meisten Befragten die ÜGK kaum eine Funktion bezüglich der Generierung von Handlungswissen für die konkrete Schulpraxis, für die Weiterentwicklung des Unterrichts oder kantonaler Bildungssysteme hat beziehungsweise haben kann. Dies vor allem, da die Ergebnisse nicht einzelnen Lehrpersonen zurückgemeldet und für die konkrete Weiterentwicklung des Unterrichts genutzt werden könnten. Zudem können im Rahmen der ÜGK-Erhebungen keine Informationen gewonnen werden, wie allfällige Probleme am besten behoben werden könnten. Während ersteres nicht als Problem angesehen wird – das Geben einer individuellen Rückmeldung an die Lehrpersonen sei nicht die Funktion der ÜGK – wird das Fehlen konkreter bildungspolitisch nutzbarer Hinweise für die Veränderung der Bildungspraxis zum Teil bemängelt. Von den Vertreter*innen der Schulpraxis wird die hauptsächliche Ausrichtung der ÜGK auf die Funktion der Systemevaluation allerdings ausdrücklich begrüsst. Darüber hinaus argumentieren viele Kantone, dass sie eigene Instrumente hätten, die explizit die Funktion der Standortbestimmung und Unterrichtsentwicklung haben. Die konzeptionell angedachte Abgrenzung zwischen nationaler Systemevaluation und kantonalen Erhebungen zur individuellen Standortbestimmung gilt somit generell als sinnvoll. Fehlen hingegen diese Instrumente in einem Kanton, so wird oftmals formuliert, dass die ÜGK mehr Funktionen übernehmen sollte.

Die positive Einschätzung der ÜGK bezüglich der verfolgten Ziele hängt von der im eigenen Kanton vorhandenen Datenqualität und -dichte ab.

Insbesondere Kantone, die selbst oder im Verbund mit anderen Kantonen Instrumente zur Erfassung der Lernleistungen der Schüler*innen implementiert haben, sprechen der ÜGK eine *ergänzende Funktion* zu, ergänzend zu den eigenen innerkantonalen und sprachregionalen Analysen und Erhebungen. Kantonale Erhebungen werden von lokalen Fachexpert*innen erarbeitet und teilweise von den Lehrpersonen korrigiert, womit sie als weniger objektiv gelten können. Aus dieser Perspektive wird die ÜGK vor allem als externes, unabhängigeres und explizit auf das vergleichende Systemmonitoring ausgerichtetes Instrument interpretiert und geschätzt. Dabei ist die interkantonal vergleichende Perspektive zentral. Diese Gruppe ist mit der jetzigen Konzeption der ÜGK als Teil einer breiteren Erhebungslandschaft mehrheitlich zufrieden und warnt vor einem Ausbau dieses Instruments. Diese Kantone werten den Umstand, dass die ÜGK anders ausgerichtet ist als PISA (andere Stichproben, Messen von Grundkompetenzen anstatt von Niveaustufen), besonders positiv und als Hinweis darauf, dass die Instrumente gut aufeinander abgestimmt sind.

Demgegenüber sprechen Kantone, welche zwar an Vergleichen interessiert sind, jedoch nicht an regionalen Leistungserhebungen teilnehmen oder selbst über keine umfassende Datenbasis verfügen, der ÜGK eine *Ersatzfunktion* zu. Diese Kantone erwarten, dass die ÜGK als vollwertiger und massgeschneiderter Ersatz für die nicht mehr stattfindenden regional-vertiefenden PISA-Analysen auf Kantonsbasis dient. Zu dieser Gruppe gehören unter anderem auch die französischsprachigen Kantone, welche die Entwicklung einer sprachregionalen Leistungserhebung nach der Einführung der ÜGK auf Eis gelegt haben, um die Schulen nicht zu überfordern.[6] Aber auch Kantone, die selbst über keine Monitoringinstrumente verfügen, können dieser Gruppe zugeordnet werden. Diese Kantone würden sich von

[6] Anstatt sprachregionaler Tests bauen diese Kantone eine regionale Aufgabendatenbank auf, die Lehrpersonen verwenden können, um den Stand ihrer Schulklassen zu evaluieren. Der Kanton Genf hat anhand der sprachregionalen Lehrpläne eigene standardisierte Leistungstests entwickelt und eingeführt.

den ÜGK eine umfassendere Ergebnisdokumentation wünschen und sehen somit die Funktion der ÜGK in einem erweiterten Sinne. Allerdings besteht in dieser Gruppe momentan eine gewisse Skepsis darüber, ob die ÜGK diese Funktion erfüllen kann. Diese Skepsis liegt einerseits am (noch) fehlenden Vertrauen in das Instrument der ÜGK, da diesem im Vergleich zu PISA eine schwächere wissenschaftliche Glaubwürdigkeit attestiert wird. Andererseits sprechen gewisse Kantone grundsätzlichere Probleme an, so die Besorgnis, dass das eigentliche Design der ÜGK – das ausschliessliche Messen von Grundkompetenzen oder das Fehlen von Daten zu bestimmten Fachbereichen oder regionaler Vergleiche – das Gewinnen von vertieftem Steuerungswissen verhindere. Hier werden entsprechende Anpassungen in Design und Analyse gefordert (vgl. dazu Kapitel 4).

ÜGK als Selbstevaluation der Kantone oder als Fremdevaluation?

Problematisiert wird vereinzelt die Frage, ob die ÜGK eine Art Selbstevaluation der Kantone sei oder ob es sich um eine Fremdevaluation handeln sollte, die vom Bund verantwortet werden müsste. Die Konzipierung der ÜGK als Fremdevaluation durch den Bund sei zentral, um den Ertrag für die Analyse der Stärken und Schwächen des schweizerischen Bildungssystems zu steigern und gegenüber dem Bund (Bezug nehmend auf Art. 62 der Bundesverfassung) Rechenschaft abzulegen. Dies würde aber auch bedeuten, dass die EDK keinen direkten Zugriff auf die Aufgabendatenbank und das Setzen der Standards mehr haben dürfte.

2.2 Nutzung der Ergebnisse für den eigenen Arbeits- und Zuständigkeitsbereich (A1.3)

Nutzen für die Bildungspolitik und die Bildungsadministration, kaum aber für die Schulpraxis.

Die ÜGK wird von den Vertreter*innen der Kantone als relevant für die Bildungspolitik und Bildungsadministration eingestuft. Für die Schulpraxis wird der Nutzen als sehr gering beurteilt. Die Resultate dienten in erster Linie der Bildungspolitik und Bildungsadministration für die Einordnung des Leistungsstands der Schüler*innen im Kanton (und in der Schweiz). Dies zeigt sich beispielsweise darin, dass im Kanton die Diskussion hauptsächlich im Dreieck Bildungsdirektor*in, Amtsvorsteher*in und kantonale Referenzperson stattfindet, nicht aber unter Einbezug der Lehrer*innenverbände.
Der mögliche Nutzen wird von einzelnen Akteur*innen in mehrerer Hinsicht beschrieben: Neben dem konkreten Einleiten von Massnahmen (z.B. Weiterbildungen oder Entwicklung eines neuen Lehrmittels) oder der Zuweisung von mehr finanziellen Mitteln für bestimmte Volksschulbereiche, wird von Vertreter*innen der Kantone auch beschrieben, dass die ÜGK zu mehr internen Diskussionen geführt habe. Der Blick werde bezüglich spezifischer Problemzonen geschärft. Manche seien zwar allenfalls bereits seit langem bekannt, aber durch die Ergebnisse der ÜGK würde deutlich, dass sie nun endlich angegangen werden müssten. Massnahmen, die schon länger geplant seien, könnten einfacher ins Feld gebracht werden. Darüber hinaus würde der Fokus der Politik auf das Bildungssystem gestärkt.
Ein möglicher Nutzen für einzelne Lehrpersonen im Sinne von der Bereitstellung von Handlungsempfehlungen oder als Kompetenz-Diagnoseinstrument wird von den Befragten nicht beschrieben, aber auch nicht problematisiert. Hingegen wird in einzelnen Gesprächen auf den indirekten Einfluss der ÜGK verwiesen, der sich beispielsweise darin zeige, dass Lehrpersonen und Schulen durch allfällige positive Resultate eine Bestätigung ihrer Arbeit in der Praxis erhielten, oder aber, falls die kantonalen Ergebnisse als problematisch beurteilt würden, Massnahmen für die Schulpraxis einleiten würden.

Aufwand-Nutzen-Verhältnis wird als ungünstig beurteilt, Mehrwert der ÜGK gegenüber anderen Erhebungen und Informationen nur teilweise wahrnehmbar.

In verschiedenen Gesprächen wird betont, dass der Aufwand der ÜGK im Verhältnis zu ihrem Nutzen gross oder zu gross sei. Zudem wird der Mehrwert, den die ÜGK im Vergleich zu anderen Daten und

Informationen im Bildungssystem erbringt, nur von einer knappen Mehrheit der kantonalen Vertreter*innen als eher gross (44 %) oder sehr gross (13 %) eingeschätzt. Des Weiteren zeigt die standardisierte Befragung, dass die kantonalen Vertreter*innen für die Identifikation von Stärken und Schwächen des Bildungssystems mehrheitlich andere Quellen und Informationen nutzen als die Ergebnisse der ÜGK. Am häufigsten sind dies wissenschaftliche Publikationen im Allgemeinen, Informationen aus Vernehmlassungen und Beratungen im Kanton oder in der Sprachregion, die Bildungsberichte Schweiz sowie Ergebnisse aus Erhebungen/Tests im eigenen Kanton. Einzig 31 % respektive 38 % der kantonalen Vertreter*innen geben an, die nationalen Berichte oder die vertiefenden Analysen zur ÜGK eher häufig beziehungsweise sehr häufig zu nutzen.
Auch für Entscheide bezüglich der Qualitätsentwicklung des Bildungswesens nutzen die Vertreter*innen der Kantone gemäss der standardisierten Befragung vor allem Informationen aus Vernehmlassungen, Beratungen im eigenen Kanton oder in der Sprachregion, wissenschaftliche Publikationen im Allgemeinen und Ergebnisse aus Erhebungen/Tests im eigenen Kanton. Ergebnisse aus den nationalen Berichten der ÜGK oder aus den vertiefenden Analysen werden einzig von circa 35 % der kantonalen Vertreter*innen eher häufig oder sehr häufig genutzt.

Verschiedene Änderungen im Konzept und der Durchführung der ÜGK könnten den Nutzen steigern.

In den Gesprächen werden verschiedene Vorschläge gemacht, wie man den Nutzen der ÜGK für die Identifikation von Stärken und Schwächen sowie das Fällen bildungspolitischer Entscheidungen steigern könne. Einzelne Kantone sind der Meinung, dass das Missverhältnis zwischen Aufwand und Ertrag nur durch differenziertere Ergebnisse und Analysen korrigiert werden könne. Um aus dieser Sicht den Nutzen zu steigern, müsse die ÜGK nicht nur Grundkompetenzen, sondern das gesamte Leistungsspektrum erfassen und es müsse eine höhere Kadenz in den Erhebungen erreicht werden. Letzteres wird allerdings aufgrund der finanziellen Belastung, des teils fehlenden politischen Willens und aufgrund der hohen Belastung der Schulen, insbesondere in kleinen Kantonen, teilweise kritisch beurteilt – in kleinen Kantonen sind jeweils alle Schulen von den Erhebungen betroffen.
In den Gesprächen werden weitere Gründe genannt, die dazu geführt hätten, dass die Ergebnisse der ÜGK bislang nur begrenzt für die Bildungsplanung und Bildungspolitik nutzbar gewesen seien. So wurden die ersten Erhebungen eher als Piloterhebungen interpretiert, deren Validität noch nicht richtig abgeschätzt werden konnte. Die ÜGK hätte medial bisher keine Wellen ausgelöst, wie dies bei der Veröffentlichung der Ergebnisse der PISA-Studie der Fall war, und sei damit in der Politik und Öffentlichkeit weniger präsent. Während einige befragte Personen diesen Umstand kritisieren, merken andere an, dass dies teilweise durchaus gewollt sei, da man ein mögliches Ranking zwischen den Kantonen verhindern wollte.
Da im Ablauf der Erhebungen (Jahrgangsstufen, Fächerauswahl) bisher keine klare Logik und auch keine langfristige Perspektive sichtbar geworden seien, bestünde zudem keine Planungssicherheit. Die Interpretation der Befunde sei schwierig und eine langfristige Entwicklung des Bildungssystems vor Ort kaum möglich. Falls Veränderungen im Konzept der ÜGK in Zukunft vorgenommen werden könnten, so wird der ÜGK auf bildungspolitischer und bildungsadministrativer Ebene ein sehr viel grösserer Nutzen zugesprochen (dies unabhängig davon, ob die Schulen eine direkte Rückmeldung erhalten oder nicht). Drei Viertel der kantonalen Vertreter*innen, die die Frage in der standardisierten Befragung beantwortet haben, geben sodann auch an, dass Änderungen im Konzept oder in der Vorgehensweise der ÜGK vorstellbar seien, die dazu führen würden, dass sie die Ergebnisse in ihrem Aufgaben- und Zuständigkeitsbereich besser nutzen könnten.

Ob aufgrund der ÜGK-Ergebnisse Massnahmen und Konsequenzen für Bildungspolitik oder Bildungspraxis abgeleitet werden, hängt von verschiedenen Faktoren ab.

Das Fällen von Entscheiden aufgrund der Ergebnisse der ÜGK scheint *abhängig vom erreichten Resultat* zu sein. Falls das Ergebnis der ÜGK für den eigenen Kanton in Abhängigkeit von den eigenen Erwartungen positiv ausfalle, führe dies zu keinen spezifischen Massnahmen, da diese Ergebnisse als Indikator für den Erfolg des eigenen Bildungssystems interpretiert werden. Falls das Ergebnis allerdings negativ ausfalle, werde dies als „politisches Desaster“ erlebt, was zu einem hohen sozialen Druck sowie zu Massnahmen in der Praxis führe. In einem Kanton hat jedoch schon die (geplante) Durchführung der ÜGK zu Massnahmen geführt, bevor überhaupt die Ergebnisse vorlagen.

Zentral ist die von den Akeur*innen *eingeschätzte Validität der Befunde*. In mehreren Gesprächen wird den PISA-Erhebungen eine ausgeprägte Glaubwürdigkeit attestiert. Sie gelten als wissenschaftlich fundiert und nuanciert, man schätzt die Bereitstellung benutzerfreundlicher Daten, die tiefergehende Analysen erlauben. Die ÜGK geniesst nicht überall dasselbe Vertrauen. Falls deren Validität als gering eingeschätzt wird, sei es aufgrund der aus ihrer Sicht mangelnden Durchführungsqualität, der Testmethoden, der Schwellenwertsetzung oder weil die sozioökonomischen Unterschiede zwischen den Kantonen nicht genügend berücksichtigt worden seien, werden die Ergebnisse nicht weiter in die Planung von Massnahmen einbezogen. Ebenso wird vor allem in der französischsprachigen Schweiz moniert, dass die Leistungstests und die Kontextfragebögen hauptsächlich in der Deutschschweiz konzipiert würden, daher für die Bildungspraxis in der französischsprachigen Schweiz weniger curricular valide seien. Dies sei teilweise in den Aufgabenstellungen ersichtlich geworden. Zudem würden sich die Lernkulturen und die Lehrpläne zwischen den Sprachregionen unterscheiden, so dass die Aussagekraft der Erhebungsinstrumente teils angezweifelt wird.

Ebenfalls wird moniert, dass der Kontextfragebogen in der Ergebnisdarstellung der ÜGK in den Berichten kaum Eingang gefunden habe und die Daten nicht (genügend) ausgewertet worden seien. So kritisieren mehrere Kantone, dass *das Potenzial der via Kontextfragebogen erhobenen Daten ungenügend ausgeschöpft werde***.** Einige Kantone beschreiben sich gar als „frustriert“, weil in ihren Augen die aus dem Kontextfragebogen gewonnenen Daten nicht in die Analysen eingeflossen sind.

Zudem sei der *Zugang zu den Daten zu kompliziert***,** *um eigene Analysen durchführen zu können,* insbesondere für Kantone ohne eigene Forschungsinstitutionen. Ausserdem fehle in den meisten Kantonen das Fachwissen wie auch die Ressourcen, um die Daten im jetzigen Format verwenden, die Ergebnisse angemessen interpretieren und in Massnahmen transferieren zu können.

Verschiedene Kantone wünschen sich, dass *auf der Basis der ÜGK tiefergehende und spezifischere Analysen erarbeitet werden*. Dieses Bedürfnis ist besonders ausgeprägt in einigen kleineren Kantonen, die selbst nicht über spezialisierte Bildungsplanungs- und Bildungsforschungsstellen verfügen. Insbesondere hier wird die Kritik laut, es könne aus den von der ÜGK erhobenen Daten zu wenig Steuerungswissen gewonnen werden, vor allem angesichts des Aufwands und der Kosten. Hier wünscht man sich die Möglichkeit, anhand der erhobenen Daten spezifischere Fragestellungen oder gezieltere Vergleiche erarbeiten zu lassen (z.B. zwischen verschiedenen städtischen Gebieten oder ähnlichen Schülerpopulationen in verschiedenen Kantonen). Momentan brauche die Erarbeitung der Vertiefungsanalysen, die die Kantone in Auftrag geben können, zu viel Zeit. Eine Nutzung der Ergebnisse für Bildungspolitik und Bildungspraxis sei daher kaum zeitnah möglich. Gewünscht wird von verschiedenen Akteur*innen, dass die vorhandenen Daten vollständiger ausgewertet werden und dass diese Analysen zum Kern der ÜGK werden und nicht zusätzlich beantragt und finanziert werden müssen.

Schwierig bezüglich des Entscheids, ob Massnahmen ergriffen werden oder nicht, ist die Beantwortung der Frage**,** *wie viele der Schüler*innen die Grundkompetenzen erreicht haben müssen* beziehungsweise wie gross die Gruppe jener Schüler*innen sein dürfe, die die Grundkompetenzen nicht erreicht hat, ohne dass dies zu Druck in der Politik oder Praxis führe. Ebenfalls wird mehrfach darauf hingewiesen, dass, auch wenn die Ergebnisse der ÜGK nachvollziehbar seien oder andere bisherige Ergebnisse bestätigten, die *Frage offenbleibe, wie denn mögliche negative Befunde erklärt werden könnten und wie entsprechenden Herausforderungen begegnet werden könne*.

Da die Lehrpersonen und Schüler*innen keine spezifischen Rückmeldungen zu den Ergebnissen der eigenen Klasse oder zu den eigenen Leistungen erhalten, *seien die ÜGK zu weit weg von deren Alltag*. Allerdings wird dies nur dann kritisiert, wenn der ÜGK diese Funktion zugeschrieben wird.
Die *Nützlichkeit der Befunde für die Bildungspolitik und -administration wird dann besonders positiv eingeschätzt, wenn die Ergebnisse der ÜGK bisherige Befunde (z.B. PISA oder interne Tests/Erhebungen) bestätigen*. Im Falle des Vergleichs mit intern erhobenen Daten erhält die ÜGK eine Art *Legitimationsfunktion*, indem sie als externes Instrument, dem man besser trauen könne, die internen Ergebnisse bestätige und somit legitimiere. Falls allerdings die ÜGK-Befunde mit den Befunden aus anderen Erhebungen nicht oder nur in Teilen korrespondierten (weil die Schüler*innen je nach Erhebung besser oder schlechter abschneiden, oder die Ergebnisse fachspezifisch sehr stark variieren), sei die Interpretation und Kommunikation der Befunde für die Bildungspolitik und -administration anspruchsvoll und für manche politischen Akteur*innen schwierig zu akzeptieren.

2.3 Umfang der Evidenz im Gesamtsystem (A2.8; A2.9)

Unklarheit, ob ÜGK ausgebaut werden soll oder im aktuellen Umfang funktional ist.

Trotz der eben angesprochenen Probleme wird im Allgemeinen nicht gewünscht, neue Erhebungsinstrumente zu etablieren, vor allem nicht auf nationaler Ebene. So besteht generell die Meinung, dass die Anzahl standardisierter Leistungsmessungen so gering wie möglich gehalten werden müsse. Standardisierte Erhebungen erforderten (zu) viel Unterrichtszeit und sollten sowohl im Schulalltag als auch in der Bildungspolitik kein übermässiges Gewicht erhalten.
Allerdings gibt es auch Stimmen, sowohl bei den Kantonen als auch bei den Vertretungen der Lehrpersonen, die sich von einem gezielten Ausbau der ÜGK wichtiges Steuerungswissen versprechen. Konkrete Anliegen beziehen sich auf die getesteten Fachbereiche, die Reichweite der Kontextfragebögen (z.B. Einbezug der Lehrpersonen, Eltern oder Schulleitungen), die gemessenen Kompetenzstufen sowie die Kadenz.

Mehr Unterstützung bei der Interpretation der Ergebnisse und der weiterführenden Analysen gewünscht.

Es besteht ein Bedürfnis nach mehr Unterstützung bei der Interpretation vorhandener Daten sowie bei weiterführenden Analysen auf Basis der Daten, vor allem mit Bezug auf die durch den Kontextfragebogen erhobenen Kontextvariablen. Einige Kantone versprechen sich mehr Handlungswissen durch die bessere Verknüpfung der ÜGK mit existierenden kantonalen Instrumenten, entweder konzeptionell oder aber in der Durchführung. Die bessere Nutzung von Synergien zur gemeinsamen Arbeit an der Verbesserung von Tests, Fragebögen und Prozessen wird auch von den aktuell in der Gestaltung der ÜGK involvierten Expert*innen als potentiell wertvoll gesehen. So erhofft man sich, beispielsweise die Testdichte zu verringern und die Schulen zu entlasten sowie tiefergehende Analysen zu ermöglichen. Verschiedene Kantone spüren eine gewisse Überlastung von Seiten der Lehrpersonen und finden es teilweise schwierig, Schulen zum Mitmachen zu animieren. Allgemein lässt sich wiederum feststellen, dass Ansichten zu dieser Frage sehr stark mit der jeweiligen kantonalen und regionalen Test- und Dateninfrastruktur zusammenhängen.

2.4 Kommunikation der Ergebnisse der ÜGK und Berichterstattung (A1.4; A1.5)

Kantonale Vertretungen fühlen sich gut über die ÜGK informiert und setzen sich mit den Ergebnissen auseinander.

Über 87 % der kantonalen Vertreter*innen geben an, gut über die ÜGK und die Ergebnisse informiert zu sein. Etwas geringer, aber immer noch im hohen Ausmass, scheinen die kantonalen Vertreter*innen über die Art und Weise der Durchführung der Erhebungen (77 %) über die Verantwortlichkeiten zur strategischen Ausrichtung der ÜGK (74 %) und über die involvierten Institutionen, die den Auftrag haben, die ÜGK durchzuführen (69 %), informiert zu sein.

Fast alle kantonalen Vertreter*innen haben sich gemäss ihren Angaben mit den nationalen Berichten auseinandergesetzt, allerdings mit unterschiedlicher Intensität: Über 80 % geben an, dies (eher) vertieft gemacht zu haben. Die weiterführenden Berichte (z.B. zur Testentwicklung oder zum Kontextfragebogen), wurden von 41 % der kantonalen Vertreter*innen (eher) vertieft analysiert. Über 80 % der Befragten haben sich auch mit zusätzlichen wissenschaftlichen Publikationen, die mit Daten der ÜGK entstanden sind, auseinandergesetzt, circa ein Fünftel (eher) vertieft.

Sowohl von kantonalen Vertreter*innen wie auch von den Personen aus der Forschung und Aus- und Weiterbildung, die angegeben haben, die Berichte genutzt zu haben, sind die allermeisten der Ansicht, dass die Qualität der Berichte gut oder sehr gut sei.

Kommunikation zwischen Wissenschaft und Politik wird teilweise als schwierig beurteilt.

Über Kontakte mit Kosta HarmoS oder ICER wird in den Gesprächen kaum berichtet. Allerdings wird von den Vertreter*innen der Kantone thematisiert, dass die Kommunikation zwischen Wissenschaft und Politik vereinfacht werden müsse; die Wissenschaftler*innen hätten häufig ihre Hypothesen im Kopf, es brauche aber eine Sensibilität für die Praxis, die oftmals zu wenig sichtbar würde. Einzelne kantonale Referenzpersonen berichten in den Gesprächen, dass sie zweimal pro Jahr an einem Treffen in Bern teilnehmen, an dem sie über die wichtigsten Punkte informiert werden. Allerdings wird der Informationsgehalt nicht als sehr gross erlebt, da die Informationen oftmals bereits über die Bildungsdirektor*innen, die direkt über die EDK in die Diskussionen involviert sind, weitergeleitet werden. Zudem wird teilweise der als „zu wissenschaftlich“ wahrgenommene Präsentationsstil kritisiert und das Gefühl angesprochen, es werde an diesen Treffen kaum diskutiert – man werde eher vor vollendete Tatsachen gestellt. In der eigenen Region hingegen bestünden enge und gute Kontakte, so auch mit den Durchführungszentren.

Verstärkung der Kommunikation der Ergebnisse durch die EDK und bessere Abstimmung mit kantonalen Interessen gewünscht.

Knapp die Hälfte der kantonalen Vertreter*innen (44 %) ist gemäss der standardisierten Befragung der Meinung, dass die Kommunikation der Ergebnisse durch die EDK verstärkt werden müsste. 26 % der kantonalen Vertreter*innen wünschen sich zudem eine intensivere Kommunikation durch die kantonalen Behörden.

In den Gesprächen wird teilweise moniert, dass die Kantone der Publikation der Ergebnisse durch die EDK „hinterherhinken“, da sie erst aktiv werden könnten, wenn die Ergebnisse durch die EDK publiziert sind. Die Kantone könnten die Kommunikation der Ergebnisse so nicht gut vorbereiten und seien zwangsläufig in einem „Krisenkommunikationsmodus“.

Unterstützung in der Kommunikation der Ergebnisse gewünscht.

Sowohl verschiedene Kantone als auch Vertreter*innen der Lehrpersonen wünschen sich von den Verantwortlichen der ÜGK (vom ICER oder den regionalen Durchführungszentren) mehr Unterstützung in der Kommunikation der Erhebungsmethodik, der dahinterliegenden Konzepte sowie der Resultate.

Unterstützung brauche man zum Beispiel dabei, um allfällige Unterschiede zwischen der ÜGK und PISA sowie deren Implikationen zu erklären.
Aufgrund der teilweise fehlenden Akzeptanz des Kontextfragebogens wird konkret auch mehr Unterstützung gewünscht, um Politik, Bevölkerung und Lehrpersonen den Zweck dieses Instruments und der darin enthaltenen Items zu erklären. So sollte diesen Gruppen beispielsweise konkret aufgezeigt werden, warum spezifische Fragen gestellt werden oder weshalb die Resultate der ÜGK denjenigen von anderen Erhebungsinstrumenten zu widersprechen scheinen.
Zudem würde es sich aus Sicht einiger Teilnehmender an den Gesprächen lohnen, geeignete Kommunikationsmittel für die Lehrpersonen zu erstellen, damit sich diese besser eingebunden fühlten sowie besser verstehen könnten, was der Nutzen der Tests sei, was sie messen, was Mindesterwartungen an die Schüler*innen seien oder wie die Testaufgaben aussähen. Lehrpersonen hätten gerne individuelle Rückmeldungen. Es brauche eine klare Kommunikation, warum dies nicht möglich sei, damit keine falschen Erwartungen entstünden. Zudem wird angeregt, massgeschneiderte Kurzberichte zu erstellen, die auch von den Lehrpersonen gelesen werden könnten. Vor allem durch einen vertiefteren Einblick in die Testaufgaben könnten Lehrpersonen für die Kompetenzerfassung sensibilisiert werden.

2.5 ÜGK und Bildungsbericht (A2.10)

ÜGK-Daten und -Ergebnisse werden als relevant für den Bildungsbericht betrachtet, allerdings besteht ein zeitliches Passungsproblem, um diese in den aktuellen Bildungsberichten zu berücksichtigen.

Die durch die ÜGK generierten Daten werden als relevant für den Bildungsbericht beschrieben, auch weil die ÜGK die einzige Studie sei, die Kantonsvergleiche möglich macht.
Der Bildungsbericht habe zur Aufgabe, zu ausgewählten bildungspolitischen Fragestellungen eine Antwort zu finden, eine Beschreibung der Schweizer Bildungslandschaft anhand spezifischer Indikatoren zu ermöglichen sowie Erkenntnisse zu gewinnen, welches die Ursachen für bestimmte Phänomene seien oder wie etwas verändert werden könne. Diese Fragen werden anhand existierender Forschung sowie, an zweiter Stelle, anhand zusätzlicher Datenanalysen beantwortet. Die ÜGK könne also direkt und indirekt zum Bildungsbericht beitragen.
Die Daten der ÜGK 2016/2017 konnten bislang vor allem aus zeitlichen Passungsproblemen nicht im Bildungsbericht berücksichtigt werden. 2018 wurde die Publikation des Bildungsberichts zwar verschoben, um eine Integration der ÜGK-Ergebnisse doch noch zu ermöglichen, jedoch standen die ÜGK-Daten und entsprechende publizierte Ergebnisse vor dem Publikationszeitpunkt des Bildungsberichts nicht zur Verfügung. Ein ähnliches Szenario werde sich 2023 wiederholen, weil die ÜGK 2020 wegen der COVID-19-Pandemie verschoben worden sei und man nicht auf ältere ÜGK-Daten zurückgreifen möchte. Daher seien die Herausgeber*innen des Bildungsberichts 2023 auf Studien von Forscher*innen auf Basis der ÜGK-Daten angewiesen, die man dann allenfalls für den Bildungsbericht nutzen könnte.

Qualität der Daten und Analysen zentral, allerdings genüge das Erhebungsdesign nur teilweise den Anforderungen.

Ob die ÜGK-Erhebungen auch relevantes Erklärungswissen generieren können, hänge von der Qualität der Daten und Analysen ab. Die Verantwortlichen äussern in dieser Hinsicht eine gewisse Skepsis. Diese hänge vor allem mit dem Design der ÜGK-Erhebung zusammen. Konkret genannt werden drei Probleme: Erstens, dass die ÜGK Grundkompetenzen statt des gesamten Kompetenzspektrums erfasse. Dies mache statistische Analysen schwierig und verhindere, Erklärungswissen zu generieren. Zweitens erschwere die Anlage der ÜGK als Querschnittsuntersuchung und das Fehlen eines zuverlässigen Erhebungsrhythmus‘ die Identifikation des Effekts der Schule und des Bildungssystems auf die Entwick-

lung der Schüler*innen. Drittens verschleiere der starke Fokus auf Schüler*innen durch den Kontextfragebogen die Sicht auf andere, gemäss der Forschung relevante potentielle Erklärungsfaktoren wie zum Beispiel Lehrpersonen, Unterrichtsprozesse oder Interaktionen zwischen den Schüler*innen.

3 Zielvorgaben der ÜGK

Die nationalen Bildungsziele der EDK wurden vor zehn Jahren für die Schulsprache, die zweite Landessprache beziehungsweise Englisch als Fremdsprache sowie für Mathematik und Naturwissenschaften verabschiedet. Sie beziehen sich auf das Ende des 4., 8. und 11. Jahres der obligatorischen Schule. Genau hier müssen die ÜGK-Erhebungen ansetzen. Ebenfalls festgelegt wurde bereits im HarmoS-Projekt, dass die Bildungsziele als Grundkompetenzen (=Mindeststandards) benannt und operationalisiert werden sollen. Dementsprechend müssen alle bisherigen und in Entwicklung befindlichen ÜGK-Tests den jeweiligen Schwellenwert fokussieren, bei dem Grundkompetenzen als „vorhanden“ bezeichnet werden können.
Die Gutachtengruppe wurde jedoch von der EDK gebeten, zu prüfen, wie sinnvoll es sei, am Testen von Grundkompetenzen im Sinne der Überprüfung des Erreichens von Mindeststandards festzuhalten. Auch die Art der Beschreibung dieser Grundkompetenzen in den Bildungsstandards sowie die Festlegung auf bestimmte Fachbereiche und Jahrgänge stehen auf dem Prüfstand.
Die anstehende Gesamtschau zur ÜGK soll und muss also gewissermassen noch einmal die Prinzipien des HarmoS-Projekts, des daraus entstandenen Konkordats und die Grundlagen ihrer Umsetzung in der ÜGK prüfen. Das vorliegende Gutachten bereitet eine solche Evaluierung der Grundlagen und der Ausgestaltung der Erhebungen vor. Die Antworten und Empfehlungen stützen sich nicht zuletzt auf die Grundposition der Gutachtengruppe, dass die ÜGK eindeutig und in erster Linie als Instrument des Bildungsmonitorings und somit der bildungspolitischen Steuerung anzusehen ist, und erst nachrangig als Datenquelle für die Forschung.

3.1 Auswahl der Fächer und Jahrgangsstufen (B1.1; B1.2)

Auswahl der Fächer und Jahrgangsstufen findet Akzeptanz.

In der standardisierten Befragung wurde gefragt: „Welche der nachfolgenden Fachbereiche sollten aus Ihrer Sicht im Rahmen der ÜGK überprüft werden?“ Deutlich wurde, dass die Schulsprache und Mathematik grösste Akzeptanz finden (vgl. Tabelle 2): 90% der kantonalen Vertreter*innen sowie etwa 70% der Personen aus Forschung und Aus- und Weiterbildung hielten es für notwendig, diese beiden Fächer zu überprüfen. Etwas niedrigere Werte, aber immer noch mit einer Zustimmung von einer Mehrheit der Befragten, ergaben sich für die zwei weiteren in der ÜGK vorgesehenen Fächergruppen Fremdsprachen und Naturwissenschaften. Interessant ist, dass den Vertreter*innen der Kantone die Fremdsprachen etwas wichtiger sind, den Personen aus Forschung und Aus- und Weiterbildung hingegen die Naturwissenschaften. Nur knapp ein Drittel der Befragten votierte für die Ergänzung des ÜGK-Programms durch ein musisches Fach oder Sport; dies waren überwiegend Personen, die in der Aus- und Weiterbildung von Lehrpersonen tätig sind.

Dieses Bild wird durch die Gespräche gestützt. Zusätzliche Fächer spielten dort mit einer Ausnahme (eine Person aus der Schulpraxis wollte die ÜGK besser von PISA abgrenzen) keine Rolle, während die Naturwissenschaften mitunter kritisch gesehen wurden – weil es mit entsprechenden Tests in der Schweiz wenig Erfahrung gebe und weil es als schwierig angesehen wird, eine Konzeption zu entwickeln, die sowohl die deutschsprachige als auch die französisch- und italienischsprachige Tradition und die naturwissenschaftlichen Einzeldisziplinen angemessen berücksichtige[7]. Eine Person votierte expli-

[7] Die Sonderstellung der Naturwissenschaften ist schon in der Formulierung von Grundkompetenzen bei HarmoS angelegt: In der Schulsprache, Fremdsprachen und Mathematik sind die Grundkompetenz-Standards nach Kompetenzbereichen gegliedert, die in Mathematik „Themenbereiche“ heissen und tatsächlich auf einer Systematik von Fachinhalten beruhen, in den Sprachen auf der Systematik des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GER). Diese Kompetenz- und Themenbereiche können prinzipiell gut in Aufgabenpakete und Tests übersetzt werden. Die Grundkompetenzen Naturwissenschaften hingegen sind nach relativ abstrakten „Handlungsaspekten“ gegliedert, weil, wie es einleitend heisst, die Festsetzung fachlicher Themen in den Kerncurricula der

zit für eine Beschränkung auf Mathematik und Schulsprache, alles andere sei „Verzettelung“. Viele Befragte betonten, dass es auf die Systematik der durchgängigen Fächerwahl über alle Kohorten hinweg im Rahmen eines langfristig geplanten Erhebungsprogramms mit angemessenem Rhythmus ankomme.

Fach / Fachgruppen	Kantonale Vertreter*innen (n = 39)	Personen aus Forschung und Aus- und Weiterbildung (n = 77)
Schulsprache	90%	73%
Mathematik	90%	71%
Naturwissenschaften	59%	64%
Fremdsprachen	74%	52%
Musik	21%	20%
Bewegung und Sport	23%	17%
Gestalten	13%	18%

Tabelle 2: Zustimmungsrate zur Überprüfung verschiedener Fächer im Rahmen der ÜGK; standardisierte Befragung;
* Keine Angaben: je 24%

Auch die von der EDK vorgenommene Auswahl der Jahrgangsstufen findet Zustimmung, allerdings mit einigen Abstrichen für die jüngeren Jahrgänge. Als „sehr sinnvoll“ bewerteten 24 % der Teilnehmenden an der standardisierten Befragung die Erhebung Ende des 4. Schuljahres nach HarmoS-Zählung (also bei circa 8-Jährigen), 41 % die Erhebung Ende des 8. Schuljahres (circa 12-Jährige) und 48 % die Erhebung Ende des 11. Schuljahres (circa 15-Jährige). Nimmt man diejenigen hinzu, welche die betreffende Erhebung als „eher sinnvoll“ akzeptieren, wächst die Zustimmung auf 50 % beziehungsweise 64 % und 67 %. Die Skeptiker*innen finden sich eher unter Personen aus Forschung und Aus- und Weiterbildung als unter kantonalen Vertreter*innen. Alternative Zeitpunkte (z.B. rund um den Schuleintritt, 6. statt 4. Schuljahr, 7. statt 8., 10. statt 11.) wurden sehr selten genannt; sie würden auch eher dann Sinn machen, wenn der Zweck der ÜGK grundlegend geändert würde, zum Beispiel in eine Diagnose individueller Lernrückstände rechtzeitig vor dem Abschluss einer Schulstufe, was aber nicht der aktuellen Zwecksetzung entspricht.
In den Gesprächen gab es ebenfalls Zweifel, ob eine Testerhebung bei 8-Jährigen realisierbar sei. Dazu wird die ÜGK 2024 mit dem Schwerpunkt 4. Schuljahr wertvolle Erfahrungen liefern; erste ermutigende Erkenntnisse erbrachte 2019 eine Machbarkeitsstudie (Angelone et al., 2019). In den beiden höheren Jahrgängen wurde vor allem aus der Schulpraxis auf das Problem hingewiesen, dass in denselben Schüler*innengruppen weitere Erhebungen durchgeführt werden, etwa in der 8. Jahrgangsstufe in den französischsprachigen Kantonen Orientierungstests und in der 11. Jahrgangsstufe in einzelnen deutschsprachigen Kantonen Stellwerk-Tests; hinzu kommt gegebenenfalls PISA. Über eine langfristige Planung und eine abgestimmte Stichprobenziehung, so die häufig formulierte Forderung, sollten Überlastungen für einzelne Gruppen vermieden werden (vgl. dazu Kapitel 4).

Sprachregionen erfolgt. Hinzu kommt, dass die sogenannte HarmoS-Validierungsstudie in den Naturwissenschaften u.a. mit aufwändigen Aufgabenformaten wie Schüler*innenexperimenten gearbeitet hat. Ein solches „Performance Assessment“ bietet Vorteile hinsichtlich Aussagekraft, Tiefe der Testanforderungen, Unterrichtsnähe und Testmotivierung, verlangt aber einen höheren Aufwand an Entwicklungsarbeit und mehr Zeit beim Testen und verkompliziert die Logistik.

3.2 Mindeststandards oder differenzierte Leistungsstufen? (B1.4; B3.13; B3.15)

Die Entwicklung eines validen Testsystems setzt im Allgemeinen voraus, dass das Konstrukt (z. B. Grundkompetenz in Mathematik oder Schulsprache auf einer bestimmten Schulstufe) in dreifacher Hinsicht spezifiziert ist:

a. Die inhaltlichen Standards sind klar definiert – wenn möglich durch die Benennung kognitiver Operationen/Leistungsmerkmale („Can-do“-Statements) – und beschrieben.
b. Die im Test verwendeten, von Expert*innen systematisch entwickelten Aufgaben lassen sich mit Anforderungsmerkmalen beschreiben, die hoch mit den Beschreibungen der inhaltlichen Standards korrespondieren.
c. Niveau-Deskriptoren („Proficiency Level Descriptors“) ermöglichen eine präzise Beschreibung von Kompetenzstufen („Proficiency Levels“), wobei die beiden zuvor genannten Aspekte (Inhaltsstandards und Testaufgabenmerkmale) kombiniert werden (vgl. Tiffin-Richards & Pant, 2017).

Im Fall der ÜGK liefern die 2011 verabschiedeten HarmoS-Dokumente Beschreibungen der inhaltlichen Standards (Komponente a), während Fachdidaktiker*innen in Zusammenarbeit mit der Aufgabendatenbank der EDK (ADB) Testaufgaben entwickelten und Beispielaufgaben vorstellten, deren Merkmale in den Berichten zur ÜGK 2016 und 2017 beschrieben wurden (Komponente b). Differenzierte Niveau-Deskriptoren (Komponente c) liegen jedoch für die Schweiz nicht vor.

Die Begrenzung der Bildungsstandards auf eine Niveaustufe (Grundkompetenz, Mindeststandard) ist weitgehend akzeptiert.

Alle schweizerischen Bildungsstandards beschränken sich auf die Beschreibung einer einzigen Niveaustufe, die „Grundkompetenz“ genannt wird und den „Mindeststandard“ darstellt, dessen Erreichen in der ÜGK erfasst werden soll. Die Grundkompetenzen nach HarmoS sind im Prinzip normativ festgelegte Niveaustufen innerhalb komplexerer Kompetenzmodelle (vgl. Ramseier, Moser, Moreau & Antonietti 2008, S. 14ff.). Die vollständigen Kompetenzmodelle werden jedoch nicht expliziert, und deshalb gibt es keine Deskriptoren, die verschiedenen Kompetenzstufen gegeneinander abgrenzen.
In der standardisierten Befragung bezeichnete nur eine sehr kleine Minderheit (6 von 141 Befragten) den Ansatz, ausschliesslich Mindeststandards zu beschreiben, als „überhaupt nicht sinnvoll“; zwei Drittel jener, die die Frage beantwortet haben (circa 30% haben keine Meinung geäussert), finden dies aber „eher sinnvoll” oder „sehr sinnvoll” (Abbildung 2). Mit etwa einem Drittel Antworten in der Kategorie „eher nicht sinnvoll“ gibt es allerdings ein gewisses Zweifel-Potenzial. Dennoch forderte in den Gesprächen niemand einen Neustart von HarmoS; die Bildungsstandards von 2011 werden weitgehend als Grundlage der ÜGK akzeptiert.

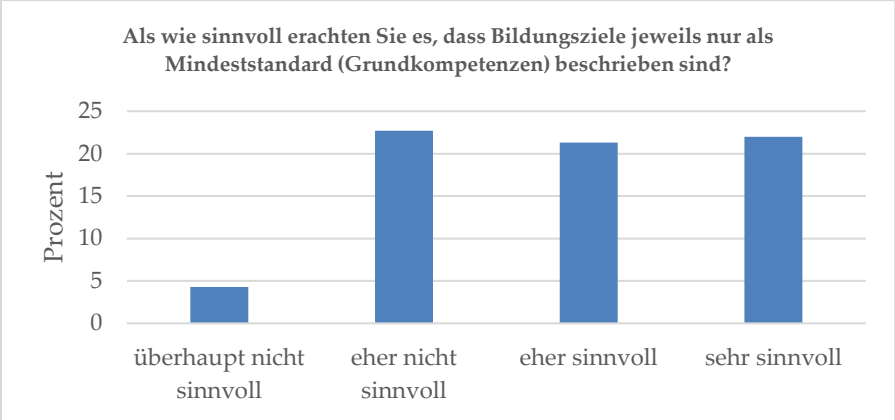


Abbildung 2: Ergebnisse standardisierte Befragung zur Fokussierung auf Mindeststandards; Prozentangaben; Keine Angaben = 29.8%

Differenzierte, das gesamte Leistungsspektrum umfassende Stufenmodelle sind umstritten.

Selbst wenn die Bildungsziele als Mindeststandards beschrieben sind, wäre es prinzipiell denkbar, bei der *Testung* im Rahmen der ÜGK ein breiteres Leistungsspektrum abzudecken und differenzierte Stufenmodelle zu formulieren, mit denen die Getesteten in hierarchische Kategorien eingeordnet werden können (z.B. „unter Mindeststandard“, „Mindeststandard“, „Regelstandard“, „über Regelstandard“ und „Optimalstandard“; vgl. Fuchs 2019, S. 7). Dies ist die Vorgehensweise in internationalen Vergleichsstudien sowie bei der Überprüfung von Bildungsstandards in Österreich und Deutschland. Dieser Ansatz wurde aber auch noch in der allerersten empirischen Studie zur Umsetzung der HarmoS-Kompetenzmodelle, der sogenannten Validierungsstudie, verwendet: Hier wurde in der Regel mit vier Stufen gearbeitet, die allerdings nicht mit einem formalen „Standard-Setting“-Verfahren abgegrenzt wurden, sodass „Methodik und Transparenz der Festlegungen insgesamt nicht befriedigend“ waren (Ramseier et al., 2008, S. 15). In der ÜGK hat man sich dann darauf beschränkt, „Grundkompetenz“ als eine einzige Stufe zu beschreiben.

Um zu klären, wie relevant die Erfassung des gesamten Leistungsspektrums ist, wurden in der standardisierten Befragung zwei Aussagen vorgelegt, welche die Teilnehmenden jeweils auf einer vierstufigen Antwortskala von „stimme überhaupt nicht zu“ bis „stimme voll und ganz zu“ bewerten sollten. Die Abbildungen 3 und 4 stellen die Ergebnisse dar, wobei circa ein Drittel der Befragten die Aussagen nicht bewertet hat.

- Bei Aussage (A) „Anhand der Tests soll einzig überprüft werden, ob die Grundkompetenzen erreicht worden sind“ teilen sich die Antworten in zwei praktisch gleich grosse Gruppen: 33 % antworten positiv, 35 % negativ.
- Der Aussage (B) „Die Tests sollen das gesamte Leistungsspektrum mit mehreren Kompetenzstufen abdecken (wie zum Beispiel bei PISA)“ stimmen 45 % der Befragten (eher) zu, insbesondere 51 % der kantonalen Vertreter*innen und 55 % der Personen aus Forschung und Aus- und Weiterbildung.

Da sich die beiden Aussagen widersprechen (jedenfalls, wenn man das HarmoS-Konzept von Grundkompetenzen voraussetzt), ist es nicht verwunderlich, dass die Antworten stark negativ korreliert sind (r=-.86). Wer Aussage A zustimmt, hat also in der Regel Aussage B abgelehnt, und umgekehrt. Die Befragten lassen sich daher zumeist als Befürwortende einer reinen Grundkompetenz-Messung oder als Befürwortende eines Stufenmodells einordnen. Das Stufenmodell hat mehr Befürwortende; die Präferenz ist aber unter Personen aus Forschung und Aus- und Weiterbildung (32:21) stärker ausgeprägt als unter kantonalen Vertreter*innen (17:13). Zugleich gibt es, vor allem unter den Personen aus Forschung und Aus- und Weiterbildung, viele mit fehlendem oder nicht eindeutigem Urteil. Befürwortet wird das Stufenmodell vor allem von jenen, die der ÜGK primär die Funktion „Systemevaluation“ zuordnen, und von jenen, die ÜGK-Berichte und nachfolgende Forschungsergebnisse selbst nutzen.

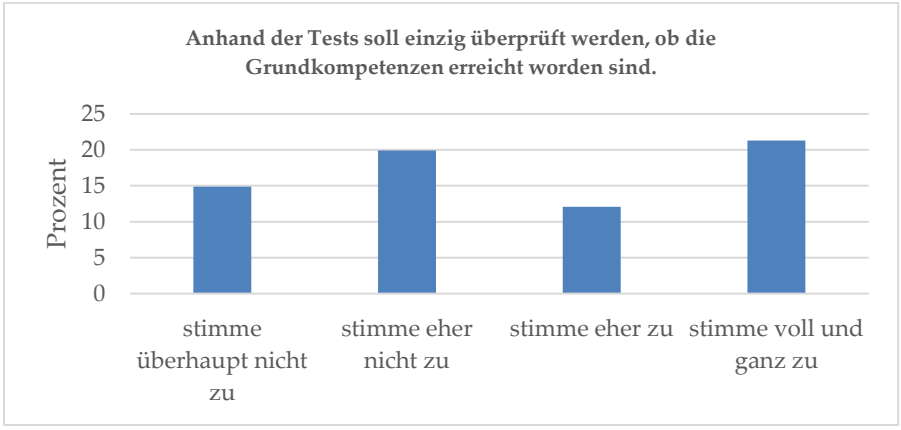


Abbildung 3: Ergebnisse standardisierte Befragung zur Frage A, ob einzig das Erreichen der Grundkompetenzen geprüft werden soll; Prozentangaben; Keine Angaben = 31.9%

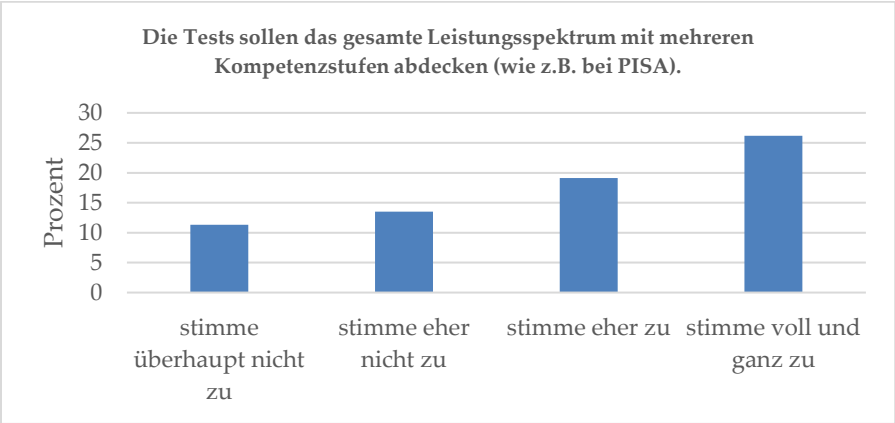


Abbildung 4: Ergebnisse standardisierte Befragung zur Frage B, Abdeckung des vollständigen Spektrums an Kompetenzstufen; Prozentangaben; Keine Angaben = 29.8%

Das Ergebnis der standardisierten Befragung ergibt somit tendenziell ein Votum für eine andersgeartete, das vollständige Spektrum umfassende, abgestufte Testskala. Auch mehrere Gesprächsteilnehmenden äusserten den Wunsch, mit der „Ressourcenverschwendung für einen 0/1-Indikator“ (Zitat aus einem Gespräch) aufzuhören und das ganze Leistungsspektrum zu testen. Diese Gesprächsteilnehmenden haben sich teils eher für eine Differenzierung unterhalb des Schwellenwerts interessiert (mit dem Anliegen, Förderanregungen zu erhalten), teils eher für Differenzierung oberhalb (mit dem Anliegen, auch etwas über leistungsstarke Schüler*innen zu erfahren). Sie erwarteten jedenfalls von differenzierten Stufenmodellen bessere Chancen, über kantonsbezogene Prozentangaben (Anteil derjenigen, die Grundkompetenzen erreicht haben) hinaus Analysen durchführen und somit Erklärungswissen aufbauen zu können. Dies führte auch zur vergleichsweise starken Befürwortung eines Stufenmodells unter Personen aus Forschung und Aus- und Weiterbildung. Einige von ihnen formulierten die Ansicht, die 0/1-Variable „Grundkompetenz nicht erreicht“/“erreicht“ sei für weiterführende Analysen und den Aufbau von Erklärungswissen ungeeignet.

In den Gesprächen wurden andererseits – vor allem von Vertreter*innen der Kantone und von Fachdidaktiker*innen – starke Argumente für die Ursprungsidee der ÜGK, also die Fokussierung auf Grundkompetenzen, vorgetragen. Eines der Argumente war hier, dass für differenzierte Messungen und Rückmeldungen in den Sprachregionen andere Instrumente verfügbar seien. Des Weiteren wurde argumentiert, dass durch den Fokus auf Mindeststandards vor allem die Leistungsschwächeren in den Blick kämen, um die sich die Schule beziehungsweise Bildungspolitik besonders kümmern müsse. Einige der Befragten waren zudem skeptisch, dass der Übergang zu einer vollständigen Skala über die Sprachgrenzen hinweg konsistent gelingen könne, weil es jenseits der Grundkompetenzen, auf die sich zu einigen schon schwer genug war, je nach Sprachregion und damit verbundener didaktischer Tradition unterschiedliche Kompetenzkonzepte und Erwartungshorizonte gebe. Schliesslich wurde darauf hingewiesen, ein Bruch mit der bisherigen Strategie impliziere einen vollständigen Neuanfang, der schon bei den 2011 veröffentlichten Beschreibungen der Grundkompetenzen ansetzen müsste. Jene Darstellungen zielten auf ein einziges Kompetenzniveau, das in den Bildungsstandards nach Themen- und Kompetenzbereichen sowie Handlungsarten aufgeschlüsselt und in den Aufgaben umgesetzt werde. Alle diese Schritte müssten erneut, ohne den Fokus auf Mindeststandards, durchlaufen werden. Ein solches Vorgehen würde grosse Kosten verursachen, sowohl die Messreihe der ÜGK als auch den politischen Prozess der Harmonisierung gleichsam auf „Start“ zurücksetzen.

Auch ohne Stufenmodelle sind Studien möglich, die über die erreicht/nicht erreicht-Klassifizierung hinausgehen.

Die nationalen (Erst-) Berichte zur ÜGK 2016 und 2017 behandeln konsequent nur das Erreichen oder Verfehlen des festgelegten Schwellenwerts, setzen dieses dichotome Ergebnis in Bezug zu anderen Variablen (z.B. Geschlecht und Migrationshintergrund) und vergleichen die Anteile der Schüler*innen, deren Testscores den Schwellenwert übersteigen, zwischen den Kantonen.
Das bedeutet aber nicht, dass Sekundäranalysen – beispielsweise für die Bildungsberichterstattung – sich ebenfalls auf eine dichotome Metrik beschränken müssen. Vielmehr liegen für alle Getesteten die Scores (= geschätzte Personen-Fähigkeiten) auf der kontinuierlichen Metrik des Rasch-Modells vor, in der Masseinheit „Logit“. In den Technischen Berichten werden die empirisch beobachteten Verteilungen als Balkendiagramme dargestellt (vgl. Abbildungen 5 und 6). Man sieht, dass insbesondere der Mathematik-Test der ÜGK 2016 und der Lesetest für Deutsch als Fremdsprache 2017 Testscore-Verteilungen ergaben, welche der Gauss’schen Normalverteilung sehr nahekommen, also für statistische Analysen bestens geeignet sind. Andere Tests aus der ÜGK 2017 haben eine leicht schiefe Verteilung, aber Deckeneffekte liegen auch hier nicht vor, sodass ausreichend Varianz abgedeckt wird. Zudem passt, wie in den Technischen Berichten dargestellt, das Rasch-Modell gut zu den beobachteten Daten. Es handelt sich also bei den ÜGK-Tests um Indikatoren für fachliche Lernergebnisse mit guten Messeigenschaften, die auch hohe Leistungsniveaus ausdifferenzieren und sehr gut in Sekundäranalysen ausgewertet werden können.
Vorbehalte gegen eine Verwendung der Testwerte in der Bildungsforschung können also höchstens mit Mängeln der Validität begründet werden. Tatsächlich präsentieren aber weder die HarmoS-„Validierungsstudie“ noch die Dokumente zu ÜGK 2016 und 2017 Daten zur Validität der Tests, abgesehen von der Modellanpassung und von Zusammenhangsmustern mit Schüler*innenmerkmalen. Es ist überfällig, empirische Daten zum Zusammenhang zwischen ÜGK-Scores einerseits, anderen Leistungsindikatoren (z.B. kantonale Tests, kommerzielle Leistungstests, PISA-Tests, Schulnoten) andererseits vorzulegen. Die EDK sollte entsprechende Untersuchungen – beispielsweise im Kontext von Pilotierungen – in Auftrag geben. Diese Untersuchungen würden dazu dienen, belastbare Hinweise auf die Validität der Testskalen zu erbringen. Zusätzliche Validitätsargumente könnten repräsentative (!) Befragungen von Lehrkräften über die Passung zu ihrem Unterricht erbringen sowie eine längsschnittliche Verknüpfung der ÜGK-Erhebungen (vgl. dazu Kapitel 4). Ein Wechsel zu neuen, breiter ausdifferenzierten Testskalen, ohne dass das bisherige Modell in solchen Studien genutzt und überprüft wurde, wäre aus Sicht der Gutachtengruppe vorschnell.
Somit ist das Argument, vertiefende Analysen (z.B. für den Bildungsbericht) verlangten eine andere Testskala, nicht stichhaltig. Zum einen stehen auch in der jetzigen Testlogik differenzierende, kontinuierliche Messwerte zur Verfügung. Zum anderen muss hervorgehoben werden, dass der Hauptzweck der ÜGK nicht solche erweiterten Analysen sind, sondern das Bildungsmonitoring bezüglich der Erreichung von fachbezogenen Grundkompetenzen. Anders als bei PISA handelt es sich bei der ÜKG nicht um allgemeine Fähigkeitsmessungen (ability, literacy), sondern um Messungen des in der Schule erworbenen Wissens und Könnens (achievement). Da sich die ÜGK unmittelbar auf nationale Bildungsstandards bezieht, kann von einer hohen curricularen Validität (Passung zu den vorgegebenen Lehrplänen und den im Unterricht behandelten Inhalten) ausgegangen werden, die für solche Tests von zentraler Bedeutung ist. Mehr noch: Gerade weil ÜGK-Tests verbindliche Grundkompetenzen erfassen, sollten sie über alle Regionen der Schweiz hinweg maximal curricular valide sein. Hätte man – wie es einige Kritiker*innen fordern – das gesamte Spektrum möglicher Kompetenzstufen aufgenommen, wären diese Aufgaben nur noch Teilen der Schüler*innenschaft zugänglich und die curriculare Validität ginge zurück.

Die Schwierigkeit der Aufgaben variiert schon jetzt breit, auch über das Niveau der Grundkompetenz hinaus.

Dass die Testscores eine ausreichende Varianz haben, liegt wiederum daran, dass die Schwierigkeiten der Aufgaben innerhalb der bislang eingesetzten ÜGK-Tests recht breit streuen. Dies ist aus den „Punktwolken“ der Abbildungen 5 und 6 ersichtlich, wo jeder Punkt eine der eingesetzten Testaufgaben darstellt, deren empirisch ermittelte Schwierigkeit rechts auf der sogenannten Logit-Skala ablesbar ist (Logit ist die Masseinheit der Schwierigkeit im verwendeten psychometrischen Ansatz, dem Rasch-Modell). Die Abbildungen – den Technischen Berichten zur ÜGK 2016 und 2017 entnommen – belegen, dass die Schwierigkeitsparameter in allen bisherigen Tests innerhalb eines beachtlichen Intervalls von circa vier Logits variieren.

Die Streubreite der tatsächlichen Aufgabenschwierigkeiten lässt sich nicht vermeiden, denn solche Schwierigkeiten „sind konstruktionsbedingt variabel. Man kann ‚denselben‘ Aufgabeninhalt so in Form einer Testaufgabe bringen, dass diese eher schwer oder eher leicht zu lösen ist“ (Rost, 2004, S. 663) [8]. Die HarmoS-Dokumente (Bildungsstandards) aus dem Jahr 2011 können nur die Art und den Inhalt der Anforderungen benennen, die zu den Grundkompetenzen zählen. Die Standards müssen in der Aufgabenentwicklung immer wieder re-interpretiert, ausdifferenziert und in konkretere Anforderungen übersetzt werden. Der exakte Schwierigkeitsgrad einer Aufgabe hängt von vielen Aufgabenmerkmalen und komplexen, nur teilweise verstandenen Antwortprozessen ab. Jene Befragten, die mit der Aufgabenentwicklung betraut waren, sind sich der Kluft zwischen Kompetenzformulierungen der Bildungsstandards einerseits und ÜGK-Aufgaben andererseits sehr bewusst. Insbesondere die Bildungsstandards in Mathematik wurden in mehreren Gesprächen als recht vage und abstrakt bezeichnet. Dies hat vermutlich zu den Problemen bei der ÜGK 2016 beigetragen, die in einem Audit von Fischbach und Ugen (2018) diskutiert wurden.

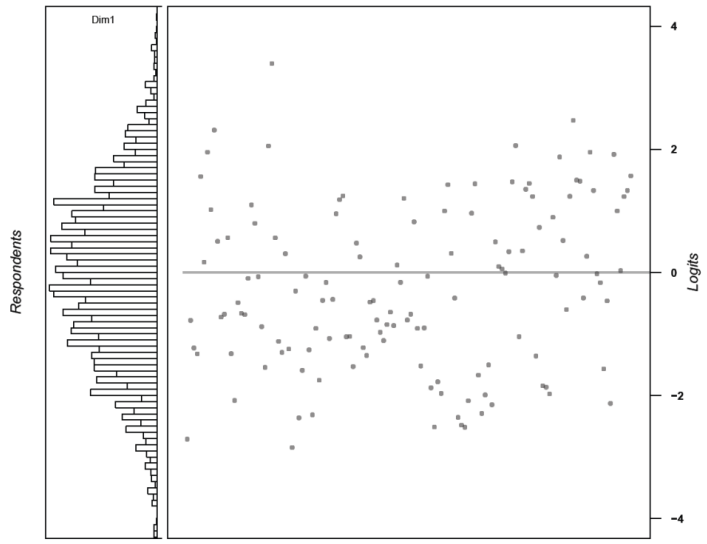
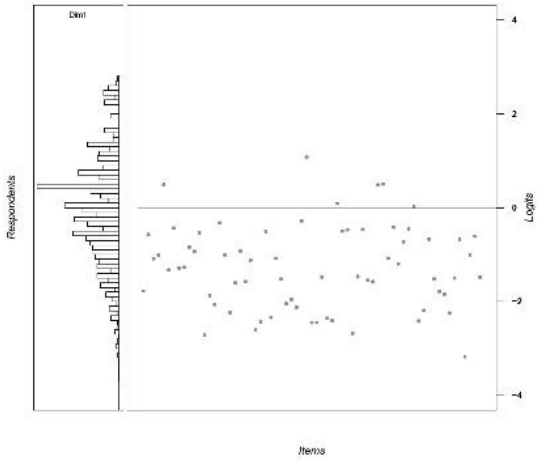


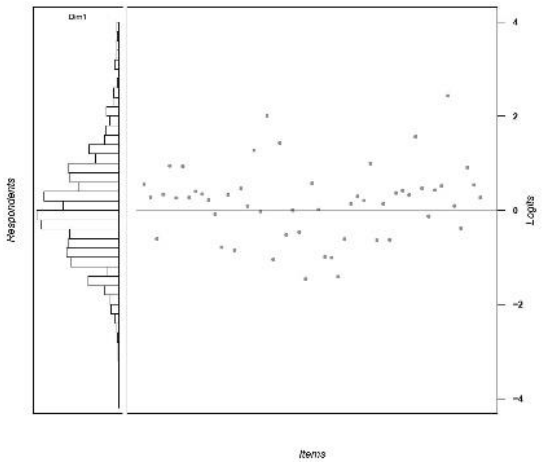
Abbildung 5: Verteilung der Personenparameter (Balkendiagramm, links) und Schwierigkeitsparameter aller Items (Punktwolke, rechts) aus der ÜGK 2016 zur Mathematik im 11. Schuljahr. Schwierigkeit eines Items wie in PISA berechnet als derjenige Score, bei welchem Probanden eine 62-prozentige Wahrscheinlichkeit haben, das entsprechende Item korrekt zu lösen (pr62). (Quelle: Angelone & Keller 2019a, S. 10)

[8] So auch das Audit von Fischbach und Ugen (2018, S. 21) zur ÜGK 2016: „ The auditors would like to emphasize that many theoretical descriptors allow for the construction of test items of very different—not to say diametrically opposed—mathematical demands and difficulties, as the empirical difficulties of the ÜGK/COFO items unmistakably testify. … The auditors' conclusion regarding the descriptors' room for interpretation is, again, not a Swiss specificity but a more general issue in the definition and evaluation of mathematics competence."

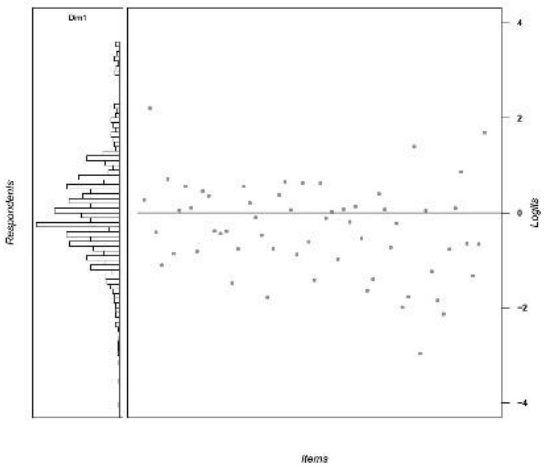
Schulsprache Lesen



Fremdsprache Deutsch Leseverstehen



Fremdsprache Deutsch Hörverstehen



Fremdsprache Englisch Hörverstehen

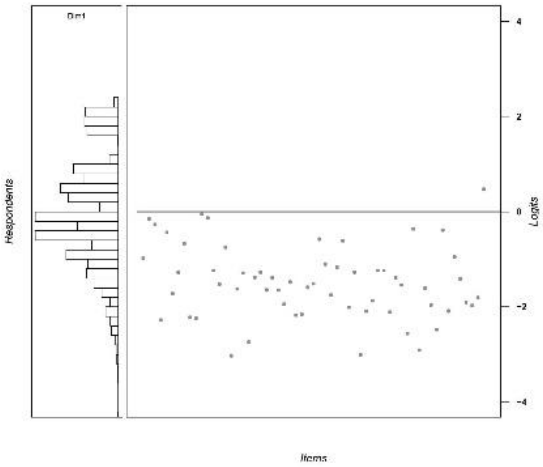


Abbildung 6: Verteilung der Personenparameter (Balkendiagramm, links) und Schwierigkeitsparameter aller Items (Punktwolke der pr62-Werte, rechts) für vier Tests aus der ÜGK 2017 zu Sprachen im 8. Schuljahr (Quelle: Angelone & Keller, 2019b, S. 17)

Wegen der unvermeidbar variierenden Schwierigkeiten muss man in der ÜGK (wie auch in allen anderen Testsystemen, die standardbasiert Bildungsmonitoring betreiben) zu dem Hilfsmittel greifen, gezielt eine gewisse Bandbreite von Aufgaben zu entwickeln, diese auf einer Testskala abzubilden und dann anhand der konkret vorliegenden Aufgaben über ein „Standard Setting“-Verfahren zu entscheiden, wo genau das angestrebte Niveau liegen soll. Dementsprechend geht der Messansatz, der der ÜGK zugrunde liegt, von einem *Kontinuum* aus, auf dem Aufgaben einerseits und Fähigkeiten der Getesteten andererseits verankert werden können. Das Rasch-Modell dient dazu, dieses Kontinuum darzustellen. Im Rahmen der Schwellenwert-Setzung („Bookmark-Verfahren“) werden die Aufgaben nach ihrer Position auf der kontinuierlichen Skala, d.h. nach ihrer Schwierigkeit, angeordnet, sodass Expert*innen eine Trennlinie ziehen können zwischen Aufgaben, die von „gerade eben grundkompetenten“ Schüler*innen noch sicher gelöst werden, und solchen, deren Anforderungen mehr oder weniger über Grundkompetenz hinausgehen[9]. Dieses Vorgehen *setzt also voraus*, dass die Tests etliche Aufgaben enthalten, die über das Anforderungsniveau der Grundkompetenz hinausgehen. Die Einbeziehung schwierigerer Aufgaben ist also keinesfalls als „Fehler“ in der Testentwicklung anzusehen!

[9]Im Technischen Bericht zur ÜGK Mathematik 11 beschreiben Angelone und Keller (2019a, S. 14) die Schwellenwertsetzung folgendermassen: „Anschliessend mussten die Experten in sprachregional zusammengesetzten Gruppen auf der Grundlage eines nach Itemschwierigkeiten gereihten Itembuchs an der Stelle ein Lesezeichen (Bookmark) setzen, an der sie die Schwelle zwischen mathema-

Allerdings liefern die Bildungsstandards selbst für solche anspruchsvolleren Aufgaben keine explizite Grundlage, denn sie beschreiben nur das Niveau der Grundkompetenz. Die fachdidaktischen Teams, die die Aufgaben entwickelt haben, sind mit dieser Herausforderung unterschiedlich umgegangen, was in den Gesprächen dargelegt wurde, aber in den offiziellen Berichten nur schwach angedeutet wird.

- Das Fremdsprachen-Team verweist darauf, dass dies nur möglich gewesen sei, indem man neben den Schweizer Bildungsstandards „die Systematik des Gemeinsamen Europäischen Referenzrahmens für das Sprachenlernen“ (GER) herangezogen habe. So habe man neben dem im EDK-Dokument beschriebenen GER-Niveau A1.2 auch Aufgaben einsetzen können, die A2 abbilden. „Der Einbezug eines Niveaus, das über den Grundkompetenzen liegt, war dadurch möglich, dass sich die Grundkompetenzen auf eine weiterführende Skala, die GER-Skala, beziehen. Der Test sollte insbesondere im kritischen Skalenbereich an der Nahtstelle zwischen den Niveaubändern A1.2 und A2.1 genügend Aufgaben aufweisen, um eine zuverlässige Grenzziehung zu ermöglichen“ (Lenz et al. 2019, S. 37f.).
- Das Mathematik-Team (Linneweber-Lammerskotten et al., 2019, S. 18) verweist darauf, dass den Bildungsstandards dieses Faches ein mehrdimensionales Kompetenzmodell mit mehreren Niveaustufen zugrunde läge, das zum Beispiel Erfahrungen aus PISA aufgreife, auch wenn dieses Modell weder in den Schweizer Bildungsstandards noch im Bericht zur ÜGK 2016 expliziert wird. Im Gespräch schlägt eine Vertretung dieser Fachgruppe vor, zukünftig für jeden ÜGK-Test ein „Framework“ zu veröffentlichen, welches die Bildungsstandards konkretisiert, neue fachdidaktische Entwicklungen aufgreift, und dann auch Beispielaufgaben sowie Aufgabenmerkmale benennt, die Anforderungen unterhalb und oberhalb der Schwelle unterscheiden.
- Das Team Schulsprache hat sich am ehesten darauf beschränkt, Aufgaben im unteren Schwierigkeitsbereich, um den möglichen Schwellenwert herum, zu entwickeln – durchaus erfolgreich, wie man in Abbildung 6 sieht. Dazu hätten, so eine interviewte Person, Hinweise von Psychometrikern beigetragen. In Vorbereitung der ÜGK 2024 sei diese Fokussierung noch besser möglich, weil man tatsächliche Schwierigkeiten und kognitive Anforderungen in den „Kognitive Labs“ prüfen könne.

Diese verschiedenen Strategien führten zu sehr unterschiedlichen Verteilungen der Aufgaben auf die Bereiche unterhalb und oberhalb der jeweils gesetzten Schwelle. Im Lesetest Schulsprache war das Verhältnis ausgewogen: 35 Aufgaben lagen darunter, 32 darüber[10]. In der ÜGK 2016 Mathematik aber gab es 45 Aufgaben unterhalb und 87 oberhalb der Schwelle, und bei den Fremdsprachen lagen beim Leseverstehen nur 11 bis 13 (je nach Sprache) von 53 Aufgaben unterhalb der Schwelle, im Hörverstehen 13 bis 21 von 59 Aufgaben[11]. Diese geringe Zahl ist tatsächlich problematisch, weil allein die Aufgaben, die unterhalb der Schwelle liegen und somit von „minimal grundkompetenten“ Schüler*innen mit mindestens 62 % Wahrscheinlichkeit gelöst werden können, das ausmachen, was empirisch mit „Grundkompetenz“ gemeint ist. Die Aufgaben oberhalb der Schwelle dienen hingegen zur Abgrenzung, denn sie stellen irgendwelche Anforderungen, die über das hinausgehen, was empirisch als „grundkompetent“ bezeichnet wird. Es geht bei der ÜGK eben *nicht* um ein „Mastery Learning“ wie in den 1960er und 1970er Jahren, wo man ein Aufgabenuniversum als Zielbereich festlegte und dafür die Erfolgswahrscheinlichkeit prüfte. Es geht vielmehr um die Abgrenzung zwischen zwei Stufen auf einem Kontinuum: „Grundkompetenz erreicht“ (Personen-Score liegt oberhalb des Schwellenwertes, sodass alle Aufgaben unterhalb der Schwelle beherrscht werden) versus „Grundkompetenz nicht erreicht“ (Personen-Score liegt unterhalb des Schwellenwertes).

tischen Grundkompetenzen und höheren Kompetenzen vermuteten.“ Diese Formulierung erscheint der Gutachtengruppe klärungsbedürftig, da nicht sicher ist, dass die Expert*innen des Standardsetting-Panels tatsächlich instruiert wurden, sich eine „*minimal* grundkompetente” Person vorzustellen.

[10] Dies lässt sich aus den im Anhang im Technischen Bericht (Angelone & Keller, 2019b) angegebenen Item-Kennwerten erschliessen, wenn man diese mit dem Schwellenwert von -1,260 Logits (ebenda, S. 20) abgleicht.

[11] Im Technischen Bericht zur ÜGK 2016 wurden diese Zahlen angegeben (Angelone & Keller 2019a, S. 14); für die Fremdsprachen lassen sie sich wiederum aus den Schwellenwerten und Item-Parametern rekonstruieren.

Aus Sicht der Gutachtengruppe ist also die Bandbreite der Tests an sich kein Problem – im Gegenteil: Um den Schwellenwert herum braucht es eine besondere Häufung von Aufgaben. Aber man braucht auch deutlich höhere und niedrigere Anforderungen, um ein Kontinuum aufzuspannen. Zum Problem wird die Breite erst, wenn zu wenige Aufgaben verbleiben, um zu definieren, was „grundkompetente“ Schüler*innen wissen und können. Probleme können auch entstehen, wenn die Logik der Vorgehensweise und die dabei entstehenden Differenzierungen intransparent bleiben. Das scheint bei den bisherigen ÜGK der Fall zu sein: Darüber, was die beiden Stufen „erreicht“ versus „nicht erreicht“ fachlich-inhaltlich unterscheidet und wie sich dies zur Definition der Grundkompetenzen in den Bildungsstandards verhält, liegen bislang weder aus den HarmoS-Dokumenten des Jahres 2011, noch aus den Pilotierungen, noch aus der psychometrischen Analyse der ÜGK-Daten oder aus der Schwellenwert-Setzung Spezifikationen vor[12]. Das erschwert nach Meinung der Gutachtengruppe die Kommunikation der Ergebnisse in die Praxis und Wissenschaft.

Die Gutachtengruppe sieht es als notwendig an, zu diesen Fragen eine transparente, fächerübergreifende Position zu entwickeln, indem Psychometriker*innen und Fachdidaktiker*innen darstellen, was im Rahmen der ÜGK unter „Messung von Grundkompetenzen“ verstanden wird, wie man also von der Beschreibung der Grundkompetenzen in den HarmoS-Dokumenten zu einer operational unterlegten und fachdidaktisch interpretierbaren Unterscheidung zwischen „erreicht“ und „nicht erreicht“ kommt. Möglicherweise würde sich dabei zeigen, dass der Begriff „Grundkompetenz“ in zwei unterschiedlichen Bedeutungen verwendet wird: Zum einen als das Insgesamt der Kompetenzen, die in den ÜGK-Tests erfasst werden, und zum anderen als die untere von zwei Stufen der Kompetenzskala (in Abgrenzung zu höheren Anforderungen); eine Klärung wäre dann von großer Bedeutung. Der Verweis auf das Verfahren der Schwellenwert-Bestimmung allein reicht nicht – es kommt darauf an, dessen Ergebnis inhaltlich verständlich zu machen, zum Beispiel exemplarisch durch Benennung von typischen Anforderungen der bisherigen Tests, die unterhalb und oberhalb der jeweiligen Schwelle lagen. Ein solches Positionspapier müsste deutlich machen, dass – trotz der Fokussierung auf Mindeststandards – Aufgaben mit einer gewissen Bandbreite benötigt werden, und dass man für eine valide Testkonstruktion Aufgabenmerkmale identifizieren muss, welche nachweislich die Schwierigkeit des Tests beeinflussen.

[12] Zu einigen Beispielaufgaben der ÜGK 2016 (Mathematik) werden im Bericht mögliche Varianten skizziert, die eine höhere Schwierigkeit hätten, aber es kommt zu keiner verallgemeinerten Differenzierung.

Es fehlt an systematischen Analysen zu Aufgabenmerkmalen.

Die an den Gesprächen beteiligten Fachdidaktiker*innen sind durchweg an solchen Aufgabenanalysen interessiert. Es hat bereits entsprechende Explorationen gegeben, diese wurden aber nicht systematisch ausgewertet und publiziert. Nach Ansicht der Gutachtengruppe sollten Aufgabenanalysen gezielt von der Politik beziehungsweise Bildungsadministration in Auftrag gegeben und budgetiert werden, denn es handelt sich hierbei nicht um Studien mit primär wissenschaftlichem Interesse, sondern um Massnahmen im Rahmen der Aufgaben- und Testentwicklung, die für eine solide Fundierung und eine transparente Kommunikation der Schwellenwerte unerlässlich sind. Die Neuerungen in der Aufgabenentwicklung, die als Antwort auf das Audit von Fischbach und Ugen (2018) zur ÜGK 2016 für 2024 umgesetzt werden, weisen den richtigen Weg: „Kognitive Labs“ mit Schüler*innen, Befragungen von Lehrpersonen und quantitative Pilotierungen können systematische Anforderungsanalysen auf den Weg bringen und somit genauere Beschreibungen dessen, was „Grundkompetenz“ ausmacht, ermöglichen.

Anhand der Messungen in Deutsch, Englisch und Französisch – jeweils als erster Fremdsprache – wird besonders deutlich, warum vertiefende Aufgaben- und Testanalysen wichtig sind. Hier wurden 2011 die Grundkompetenzen in den Bereichen Leseverstehen und Hörverstehen für alle drei Sprachen identisch formuliert und bei der ÜGK 2017 in einen einheitlichen Aufgabensatz übertragen. Empirisch verhielten sich die Aufgaben in den drei Sprachen, die jeweils in unterschiedlichen Kantonen als erste Fremdsprache geprüft wurden, jedoch durchaus verschieden, wie Abbildung 6 zeigt: Dieselben Testaufgaben zum Hörverstehen sind nämlich, gemessen an der jeweiligen Kompetenzverteilung der Schüler*innen (vgl. Häufigkeitsdiagramme im jeweils linken Teil der Grafiken), in der Fremdsprache Deutsch schwieriger als in der Fremdsprache Englisch. Darauf hat man bei der Schwellenwert-Setzung offenbar mit unterschiedlich anspruchsvollen Grenzwerten reagiert: Von den 59 identischen Aufgaben beherrscht ein „grundkompetenter“ Schüler oder eine „grundkompetente“ Schülerin im Englischen 21, im Deutschen aber nur 13 Aufgaben[13]. Die Problematik der etwas unterschiedlichen Testnormen trotz identisch formulierter Mindeststandards wurde im Bericht nicht erwähnt. Damit wurde eine Chance vertan, die Schwellensetzungen – auch mit der ihnen innewohnenden gewissen Willkür – transparent zu machen, sie fachdidaktisch zu interpretieren und wichtige Erkenntnisse über die Realität des Sprachraums Schweiz zu generieren. Die Daten der ÜGK 2017 könnten ferner über sogenannte DIF-Analysen (Differential Item Functioning) dazu genutzt werden, diese einmalige Realität besser zu verstehen, differenzierte Hinweise für den Sprachunterricht zu gewinnen und zu reflektieren, was dieser Sprachkontext für die Harmonisierung von Bildungszielen und -ergebnissen bedeutet[14]. Damit könnte das Anliegen von HarmoS konzeptuell und empirisch stark vorangebracht und mit dem fachdidaktischen Praxisdiskurs verknüpft werden.

[13] Dies ergibt sich wiederum, wenn man im Tabellenanhang zum Technischen Bericht die p62-Werte der Aufgaben mit dem jeweiligen Schwellenwert vergleicht. Die Schwelle liegt für Hörverstehen Englisch bei -1,767 Logits, für Hörverstehen Deutsch bei -0,877 Logits; der Nullpunkt der Logit-Skalen entspricht, wie in den Abbildungen 5 und 6 dargestellt, dem jeweiligen Mittelwert der getesteten Schülerpopulation.

[14] Für erste Überlegungen in dieser Richtung vgl. Schneider, Lenz, Studer (2009, S. 35)

4 Design der ÜGK

4.1 Erhebungsrhythmus und Stichprobenziehung (B1.3; B1.5; B1.6; B2.8)

Systematische und längerfristige Planung für den Erhebungsrhythmus wird dringend gewünscht.

Drei Zitate aus den offenen Schlussbemerkungen, um die in der standardisierten Befragung gebeten wurde, können die Position beschreiben, die mehrheitlich auch in den Gesprächen eingenommen wurde:

- „Die Frage nach den Testzeitpunkten erscheint mir zentral.“
- „Mir fehlt ein klare inhaltliche, terminlich langfristige (8-10 Jahre!) Konzeption, welche Fächer in welchen Schuljahren bezüglich Erreichung der Grundkompetenzen überprüft werden. Aktuell stelle ich beim Vorgehen seit Beginn eine gewisse Beliebigkeit fest.“
- „Es muss eine längerfristige, klarere Planung geben, welche Fächer für welche Schuljahre und wann. Es könnte sehr interessant sein, etwas im Längsschnitt zu planen (dieselben Schüler*innen werden in den Klassen 4, 8 und 11 in denselben Fächern getestet).“

Auch wenn sich die Vorstellungen darüber, was genau der optimale Erhebungsrhythmus sei, unterscheiden, stimmen die Befragten und Teilnehmenden an den Gesprächen darin überein, dass die EDK von einer ad-hoc-Ausschreibung einzelner Studien zu einer mittel- bis langfristigen Planung übergehen sollte. Dies hätte folgende Vorteile:

- Ein langfristiger Planungshorizont gibt allen Akteur*innen eine hohe Sicherheit: Den Kantonen und der EDK, die von immer wieder neuen Debatten über Einzelerhebungen verschont werden sowie auch den beauftragten Wissenschafter*innen. Dahingehende Wünsche wurden in den Gesprächen vor allem von der Projektleitung ÜGK geäussert.
- Regelmässige Erhebungen zum selben Fach in derselben Altersgruppe erlauben Trendaussagen, die eher geeignet sind, Bildungspolitik und -administration zu informieren[15].
- Allen Beteiligten, auch der Öffentlichkeit, würde klar, dass die ÜGK-Erhebungen keine einmaligen Schnappschüsse sind, sondern von der regelmässigen Wiederholung leben, die es erst ermöglicht, Aussagen über Trends zu machen. Das könnte die Aufregung um einzelne Erhebungen sowie die Fixierung der öffentlichen Debatte auf Rankings, die immer wieder kritisiert wurde, reduzieren.
- Man könnte klarer definieren, welches die Indikatoren für „Harmonisierung“ sind. Etwa folgendermassen: Harmonisierung ist kein Zustand, sondern ein *Prozess* der Angleichung, dessen Fortschritt daran ablesbar ist, ob die Kantone sich hinsichtlich (a) des Anteils der Schüler*innen, die mindestens Grundkompetenzen erreichen beziehungsweise (b) der entsprechenden Anteile in bestimmten Teilgruppen (nach Geschlecht, Zuwanderungshintergrund, sozialem Status und allenfalls Bildungsgang) *annähern*.

Die Vorteile von Trendmessungen sind unstrittig. Vielfach wird Trendmessung (d.h. die Wiederholung der Erhebung im selben Fach und in derselben Jahrgangsstufe nach einigen Jahren), um systemische Entwicklungen abzubilden, sogar als definitorischer Kern des Bildungsmonitorings verstanden. Umstrittener sind hingegen Kohortenlängsschnitte, also die wiederholte Testung eines Jahrgangs von Schüler*innen mit jeweils neuen Stichproben, oder gar individuelle Längsschnitte – die Erfassung der Bildungsverläufe einer Stichprobe etwa von Jahrgangsstufe 4 über 8 bis 11. Beide sind ungleich aufwändi-

[15] Aus internationalen Schulleistungsstudien und regionalen Vergleichen anderer Länder kann man lernen, dass es keine wirklich überzeugenden empirischen Erklärungen für das Ausmass der zu Beginn eines Erhebungsprogramms festgestellten Leistungsunterschiede zwischen den Bildungssystemen gibt. Dies liegt an der Vielzahl der Einflussfaktoren: zum Beispiel sozio-ökonomische und kulturelle Faktoren, Schulstrukturen und pädagogische Kulturen. Eher lassen sich unterschiedliche *Veränderungen* auf Systemebene erklären, insbesondere wenn Daten zu parallelen beziehungsweise vorgeschalteten Änderungen der Bedingungsfaktoren (z.B. Änderungen der sozialen Zusammensetzung der Schüler*innenschaft, Änderung der Schulstruktur oder der Stundentafel, grundlegende Änderungen der Lehrer*innenbildung) vorliegen. Daher sind Trendaussagen unerlässlich, wenn man Anregungen für die Bildungspolitik und für administrative Massnahmen erhalten möchte. Hinzu kommt, dass Trendaussagen – anders als Aussagen über den Anteil der Schüler*innen, welche eine bestimmte Schwelle überschreiten – nicht der Willkür unterliegen, wie hoch der Erwartungshorizont gesetzt wird.

ger und setzen eine entsprechende Koppelung der Erhebungen voraus, bei individuellen Längsschnitten auch eine Koppelung der Stichproben sowie das personenscharfe Zusammenspiel der Daten. Eine solche Erweiterung wurde fünfmal in den Gesprächen gefordert sowie sechsmal in den offenen Bemerkungen zur standardisierten Befragung. Erhofft werden erweiterte Analysen zu den individuellen, familialen und schulischen Bedingungen erfolgreicher und problematischer Bildungsverläufe – vor allem, wenn zusätzlich Klassenstichproben eingeführt werden. Die Gutachtengruppe sieht hier jedoch die Gefahr einer operativen Überforderung der ÜGK und einer Überlastung auf Seiten der betroffenen Schüler*innen und Schulen. Die Gutachtengruppe empfiehlt, das ÜGK-Erhebungsdesign im Kern als Trenddesign zu konzipieren. Fraglich bleibt, in welchen Abständen Erhebungen wiederholt werden sollen.

Engere Kadenzen zwischen den Erhebungen gewünscht, allerdings könnte dies zu Überforderungen im System führen.

Wiederholt wurden in den Gesprächen engere Kadenzen gefordert. Im Detail variierten die Vorschläge zwischen einem Rhythmus von drei Jahren, einem solchen von drei bis fünf Jahren oder sechs bis neun Jahren bis zur Aussage „alle neun Jahre reicht“. Spielt man die Vorschläge konkreter durch, werden schnell Grenzen deutlich: Wenn man alle Sprachen in einer Erhebung zusammenfasst, gibt es bei der ÜGK prinzipiell je eine Erhebung zur Mathematik und zu den Sprachen in den Jahrgangsstufen 4, 8, 11 sowie zwei Erhebungen zu Naturwissenschaften, also insgesamt acht Erhebungen. Bei einem Wiederholungsrhythmus von drei Jahren müsste man also meistens drei Erhebungen pro Jahr durchführen. Bei einem Wiederholungsrhythmus von neun Jahren käme man auf etwa eine Erhebung pro Jahr. Den Vorstellungen der Projektleitung ÜGK zufolge wäre letztgenanntes mit den in der Schweiz vorhandenen personellen Ressourcen machbar, die erstgenannte Variante aber nicht.
Aus Sicht der Schulforschung und der Governance-Forschung ist ein Zeitraum von neun Jahren allerdings zu lang, um systemische Veränderungen nachvollziehen und rechtzeitig politisch reagieren zu können. Besser geeignet wäre ein Abstand von sechs Jahren zwischen zwei Trend-Messungen. Deshalb werden hier zwei mögliche Planungsmodelle für einen Zyklus von sechs Jahren vorgelegt (vgl. Abbildung 7). Beide Modelle gehen davon aus, dass jährlich eine ÜGK-Erhebung durchgeführt werden kann. Entsprechend der Befunde zum Thema „Fächerwahl“ (vgl. Kapitel 3.1) stehen Schulsprache und Mathematik im Mittelpunkt, weil eine ebenbürtige Berücksichtigung von Fremdsprachen und Naturwissenschaften entweder zu einem unrealistischen Erhebungsaufwand geführt oder die Kadenz der Erhebungen sehr erhöht hätte. Denkbar wäre aber, dass Fremdsprachen gemeinsam mit der Schulsprache erhoben werden und Naturwissenschaften gemeinsam mit Mathematik (wie z.B. in Deutschland üblich).

- Modell 1 erlaubt eine Trendmessung jeweils im Abstand von sechs Jahren. Im Abstand von drei Jahren wird jeweils dieselbe Schulstufe untersucht, allerdings mit wechselnden Fachinhalten. Das Modell hat den Vorzug, dass die Belastungen vor Ort gering sind: Jede*r einzelne Schüler*in muss höchstens einmal in seiner/ihrer Schullaufbahn an einer ÜGK-Erhebung teilnehmen; Lehrpersonen sind frühestens nach drei Jahren erneut betroffen – es sei denn, sie unterrichten sowohl in Jahrgangsstufe 4 als auch in 8.
- Modell 2 erlaubt ebenfalls eine Trendmessung im Abstand von konstant sechs Jahren sowie zusätzlich Kohortenlängsschnitte zwischen Jahrgangsstufe 4 und 11 für Sprachen sowie 8 und 11 für Mathematik, falls solche gewünscht sind.

Beide Modelle können durch entsprechende Stichprobenziehung so gestaltet werden, dass zusätzlich Längsschnitt-Analysen auf Schulebene möglich werden. Hierzu müssten Trenderhebungen bei zwei der Untersuchungen (eine an Primarschulen, eine im Sekundarbereich – in Modell 2 beispielsweise M4 und S11) eine durchgehende Kernstichprobe von Schulen umfassen. Solche Veränderungsanalysen auf Schulebene (bei denen die einzelnen Schulen selbstverständlich anonym bleiben) könnten für die Schulentwicklung besonders wichtige Informationen liefern, wenn zusätzlich schulbezogene Daten von den Schulleitungen erfragt werden.

Diese Modelle sollen nur die prinzipiellen Möglichkeiten einer langfristigen Planung illustrieren. Eine konkrete Planung muss effektive Jahreszahlen einsetzen, dabei den Anschluss an bereits durchgeführte und geplante Erhebungen, mögliche Erweiterungen des Programms (vor allem durch Naturwissenschaften) sowie Verknüpfungen mit anderen Tests (z.B. PISA, regionale Erhebungen) definieren. Die Gutachtengruppe konnte sich bereits davon überzeugen, dass die Projektleitung ÜGK auf derartige Planungen vorbereitet ist.

	Jahr 1	Jahr 2	Jahr 3	Jahr 4	Jahr 5	Jahr 6	Zyklus beginnt neu wie bei Jahr 1	
Modell 1	S4			M4			S4	
		S11			M11			S11
			S8			M8		

	Jahr 1	Jahr 2	Jahr 3	Jahr 4	Jahr 5	Jahr 6	Zyklus beginnt neu wie bei Jahr 1, aber mit längsschnittlichen Links	
Modell 2	S4 →				S8 →			S11
		S11		M4				
			M8 →			M11		
							S4 →	

Abbildung 7: Planungsmodelle für ÜGK-Untersuchungen mit Trendmessungen sowie (in Modell 2) Kohorten-Längsschnitten; S = Sprachen (Schulsprache, gegebenenfalls mit Fremdsprache), M = Mathematik, eventuell mit Naturwissenschaften, 4 = Jahrgangsstufe 4, 8 = Jahrgangsstufe 8, 11 = Jahrgangsstufe 11, Pfeile = Längsschnitt einer Kohorte.

Eine Koppelung zwischen den ÜGK-Aufgaben und den Leistungstests in den Kantonen ist wissenschaftlich komplex und aktuell ohne langfristige Perspektive weder konzeptionell noch personell zu bewältigen.

Der ursprünglich zusätzlich vorgesehene Aufgabenbereich der ADB, eine Koppelung der Aufgaben aus der ÜGK mit kantonalen Erhebungen zu unterstützen, wird von den Vertretungen verschiedener Kantone unterschiedlich gesehen. Dies hängt insbesondere davon ab, ob die jeweiligen Kantone ähnliche Daten selbst generieren oder nicht.

Grundsätzlich ist festzuhalten, dass es sich bei diesem Aufgabenbereich um eine hochkomplexe Aufgabe handelt – technisch, methodisch und insbesondere unter den Rahmenbedingungen in der Schweiz auch organisatorisch. Um diesen Aufgabenbereich konzipieren und langfristig aufbauen zu können, bedarf es jedenfalls einer langfristigen Planung, die von einem verbindlichen längerfristigen Gesamtkonzept für die ÜGK und entsprechendem Commitment zur langfristigen Implementierung abhängig ist. Unter den aktuellen Rahmenbedingungen ist dieser Teil des ursprünglich vorgesehenen Funktionsumfangs nicht zu verwirklichen. Soll dies dennoch angegangen werden, würde es dafür neben einem langfristigen verbindlichen Plan für die ÜGK-Erhebungen eine entsprechend grössere personelle Ausstattung der ADB, eine bessere Einbindung der ADB in die Gesamtstruktur der ÜGK-Durchführung sowie eine institutionelle wissenschaftliche Anbindung der ADB benötigen.

Kein Änderungsbedarf bei Standardisierung zwischen den Kantonen

Zu den Fragen, die der Gutachtengruppe gestellt wurden, gehört die folgende: „Müssen die Erhebungen zur ÜGK in allen Kantonen zwingend in derselben Zeitperiode und jeweils in denselben Fächern durchgeführt werden?“ Die Antwort muss aus wissenschaftlicher Sicht ganz eindeutig „ja“ lauten. Flexiblere Modelle würden die Vergleichbarkeit zwischen den Kantonen gefährden, ohne die das Anliegen, den Grad der Harmonisierung zu erfassen, nicht eingelöst werden kann. Auch vertiefende Auswertungen würden an Aussagekraft verlieren. Es ist aus der internationalen Forschung über Schulleistungsmessungen hinlänglich bekannt, dass Ergebnisse mit dem Erhebungsfenster zusammenhängen (also der Position innerhalb des Schuljahres und dem genauen Alter der Proband*innen) und von Kohorte zu Kohorte schwanken können. Zudem wäre die Geheimhaltung der Testunterlagen gefährdet, wenn diese teils erst mit Verzögerung genutzt werden könnten. Dieser Position haben auch die in den Gesprächen befragten Akteur*innen einhellig zugestimmt.

Kein Änderungsbedarf bei Stichprobenziehung

Eine weitere, auf den ersten Blick „technische“ Frage betrifft die mögliche Erhebung in Klassenverbänden statt mit Schüler*innen-Stichproben, die in jenen Kantonen, in denen keine Vollerhebung stattfindet, quer zu den Klassen gezogen werden. Weder in der standardisierten Befragung noch in den Gesprächen wurde ein merklicher Bedarf nach solchen Veränderungen sichtbar. Allerdings haben drei an der standardisierten Befragung Teilnehmende sowie einige an den Gesprächen beteiligte Bildungsforscher*innen angemerkt, dass Klassenstichproben notwendig seien, wenn (a) Rückmeldungen an Lehrpersonen gegeben werden sollen als Einstieg in Unterrichtsentwicklung, oder/und (b) Analysen zum Zusammenhang zwischen Klassenkontext und Schüler*innenleistung erwünscht sind. Aus der Bildungsforschung ist bekannt, dass Befunde verzerrt werden können, wenn Analysen nicht auf der angemessenen Ebene stattfinden. Beispielsweise wird PISA oft kritisiert, weil das Design keine klassenbezogenen Analysen zulässt und somit keine zuverlässigen Aussagen über Korrelate und Effekte von Unterrichtsmerkmalen machen kann. Eine Entscheidung in dieser Sache wird also davon abhängen müssen, ob die Zielsetzung der ÜGK nach wie vor allein das Bildungsmonitoring auf kantonaler Ebene ist, oder ob Auswertungen beziehungsweise Analysen für die Unterrichtspraxis erwünscht sind. Zudem spielen logistische und Ressourcen-Argumente eine Rolle: Klassenstichproben würden mehr Proband*innen erfordern.

4.2 Konzeption und Entwicklung der Kontextfragebögen (B3.16)

Stellenwert des Kontextfragebogens ist über die Jahre gewachsen, allerdings keine einheitliche Meinung, ob diese Analysen Teil der ÜGK sein sollen.

Mehreren Gesprächen zufolge ist der Stellenwert des Kontextfragebogens gegenüber den ersten Planungen der ÜGK, die von einem reinen Leistungstest ausgingen, in den letzten Jahren gewachsen. So wird die Rolle des sozialen Hintergrunds, der in Schüler*innenfragebögen erfasst wird, für bildungspolitische Diskussionen, speziell zum Thema Bildungsgerechtigkeit, von Forschenden ebenso wie von Vertreter*innen der Kantone als relevanter Bestandteil der ÜGK betont, und einzelne Aufgabenentwickler*innen sind an einem besseren Abgleich zwischen Aufgabenentwicklung und Kontextfragebogen interessiert, um Lernbedingungen untersuchen zu können.
Wie bedeutsam der Kontextfragebogen ist, wird in den Gesprächen allerdings unterschiedlich beurteilt. Ein grosser Teil der Akteur*innen ist der Meinung, dass mögliche Bedingungen des Erreichens der Grundkompetenzen nicht ohne eine differenzierte Analyse der individuellen Voraussetzungen der Schüler*innen und somit der Zusammensetzung der Schüler*innen in einem Kanton beurteilt werden könnten. Die Informationen aus dem Kontextfragebogen seien fast wichtiger als die Leistungsdaten, da

sie es erst ermöglichten, die Leistungsergebnisse angemessen zu interpretieren. Andere wiederum argumentieren, dass dies nicht Aufgabe der ÜGK sei, sondern innerkantonal geleistet werden könne und die Hauptaufgabe der ÜGK die Überprüfung des Erreichens der Grundkompetenzen der Schüler*innen sei. Dritte regen an, dass der Kontextfragebogen auch dazu genutzt werden könnte, ähnliche Kantone bezüglich der Zusammensetzung der Schüler*innenschaft zu identifizieren, da Unterschiede zwischen den Kantonen bezüglich der Zusammensetzung der Schüler*innenschaft bestünden und es keinen Sinn mache, alle Kantone mit allen zu vergleichen, sondern nur jene, die relativ homogene Bedingungen bezüglich der Zusammensetzung der Schüler*innen aufwiesen.

Hinweise zur Relevanz der Kontextfragebögen geben die Ergebnisse der standardisierten Befragung. Die Befragten sollten angeben, wie wichtig die verschiedenen Informationen für ihren jeweiligen Aufgaben- und Zuständigkeitsbereich sind. Dabei wurden Leistungsinformationen, Informationen zum sozialen Hintergrund und Informationen zur Lernumgebung unterschieden. Alle drei Punkte sahen in beiden Gruppen mehr als 60 % der Befragten als wichtig bis sehr wichtig an. Dabei unterschieden sich jedoch Personen aus Forschung und Aus- und Weiterbildung von kantonalen Vertreter*innen. Dies wird besonders deutlich, wenn man auf die Kategorie „sehr wichtig“ schaut (vgl. Abbildung 8): Für Personen aus Forschung und Aus- und Weiterbildung stellen Merkmale der Lernumgebung (z.B. des Unterrichts) eine wichtige Information dar; bei den kantonalen Vertretungen sind sie weniger relevant. Im Gegensatz dazu sind Informationen zur Herkunft der Schüler*innen für mehr als die Hälfte der kantonalen Vertretungen sehr wichtig, während dies nur für etwa 40% der Personen aus Forschung und Aus- und Weiterbildung gilt.

In den Gesprächen gab es dementsprechend sehr unterschiedliche Positionierungen zu Umfang und Weiterentwicklung der Kontextfragebögen: Während einige Praktiker*innen und Vertreter*innen der Bildungsadministration für eine Kürzung votierten, verlangten Forschende oftmals mehr Systematik in der Fragebogenentwicklung, Chancen zur Einbettung analytisch wichtiger Fragen aus der Bildungsforschung sowie auch eine Ergänzung durch Befragungen von Schulleitungen, Lehrkräften oder/und Eltern.

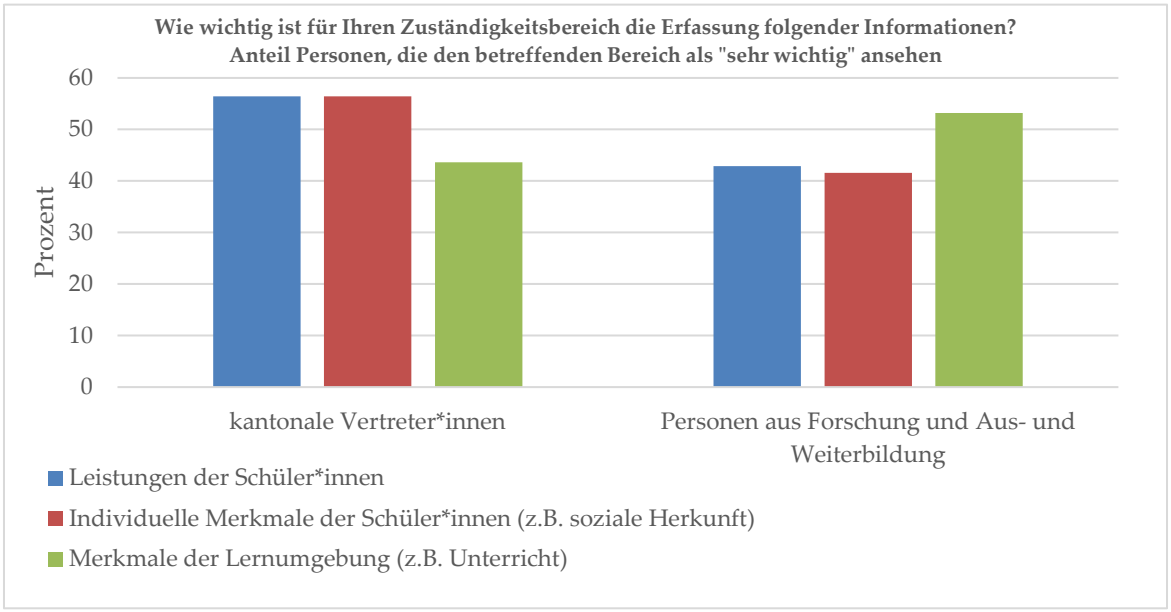


Abbildung 8: Ergebnisse standardisierte Befragung zur Bedeutung verschiedener Arten von Daten; prozentualer Anteil der jeweiligen Gruppe

Inhaltliche Rahmung des Kontextfragebogens wird kritisiert.

Eine professionelle Entwicklung des Kontextfragebogens sollte primär von vorgegebenen Fragestellungen für das Bildungsmonitoring und für vertiefende Forschung ausgehen. Prinzipiell wird der Kontextfragebogen im Large Scale Assessment dazu gebraucht, folgende Arten von Fragen zu beantworten (Kuger, Klieme, Jude & Kaplan, 2016):

1. Wie gross sind Leistungsunterschiede nach Geschlecht, sozialer Herkunft und Migrationshintergrund?
2. Welche „nichtkognitiven“ Lernergebnisse und schulbezogenen Dispositionen – etwa in Bezug auf fachliches Interesse, Lernmotivation, Selbstwirksamkeitserwartungen, Bildungsaspirationen – haben die Schüler*innen erreicht?
3. Wie hängen diese Dispositionen mit fachlichen Leistungen und dem sozialen Hintergrund zusammen?
4. Wie nehmen die Schüler*innen die Lernumgebungen und die erlebte Unterstützung in Elternhaus und Schule wahr?
5. Welchen Effekt haben die Lernumgebungen und die erlebte Unterstützung sowie Merkmale der Schule auf den Leistungszuwachs und die Entwicklung von Dispositionen?

Welche dieser Fragen in einer bestimmten Untersuchung beantwortbar sind und welche nicht, hängt wesentlich vom Stichprobendesign ab. Ein*e Gesprächsteilnehmer*in sagte dazu, Fragestellungen zur Rolle von Interaktionen zwischen den Schüler*innen könnten beispielsweise derzeit in der ÜGK nicht analysiert werden, weil keine Klassen beobachtet würden. Bevor man Themenfelder festlegt, muss daher klar sein, welches Design verfügbar ist. Ohne Klassenstichprobe können zum Beispiel für Merkmale des Unterrichts nur kantonsweite Verteilungen angegeben werden (Frage 4), aber keine sinnvollen Beziehungen zu Leistungen oder Dispositionen der Schüler*innen hergestellt werden (Frage 5). Mehrere Befragte monierten zudem, dass die ÜGK hinsichtlich der Konzeption von „Kontext“ sehr Schüler*innen-zentriert sei. Damit seien, fast zwangsläufig, am Schluss immer die Schüler*innen an den Resultaten „schuld“, weil man alle anderen potenziellen Erklärungsfaktoren ignoriere. Wie in anderen Ländern (z.B. Deutschland) könnten Eltern-, Lehrpersonen- und/oder Schulleitungsfragebögen zur Kontextdatenerhebung dazugehören, ggfs. als Option, die von Kantonen an- oder abgewählt werden kann.

4.3 Prozesse der Aufgaben- und Testentwicklung (B3.11; B3.12)

Empfehlungen des Audits von Fischbach und Ugen (2018) bei den Prozessen der Aufgaben- und Testentwicklung nur teilweise umgesetzt.

Der Audit-Report von Fischbach und Ugen (2018) kommt zusammenfassend zu Empfehlungen, die zur Verbesserung der im Audit detailliert festgestellten kritischen Aspekte der Testentwicklung im Rahmen der ÜGK vorgeschlagen wurden. Diese wurden im EDK-Bericht zu den Erhebungen 2016 und 2017 vom 21.05.2019 (EDK 2019) aufgegriffen und gliedern sich in folgende vier Hauptpunkte mit Unteraspekten[16]:

1. Die Aufgabenentwicklung aufzuwerten und zu überarbeiten, und hierbei
   - die Ressourcen für die Aufgabenentwicklung grundlegend zu überdenken
   - die Aufgabenentwicklung enger an die Testentwicklung und Datenanalyse anzubinden
   - das Koordinationsmandat der Aufgabendatenbank zu stärken
   - die Aufgabenentwicklungsgruppen um Lehrpersonen und anerkannte Assessment-Experten zu ergänzen
   - alle Aufgaben gründlich zu pilotieren
   - in rigorose Übersetzungsprozesse zu investieren

[16] Es wird hier die deutsche Übersetzung (EDK 2019) der Formulierungen von Fischbach und Ugen (2018) verwendet.

2. Die Grundkompetenzen in Mathematik zu überdenken
3. Die Leitung und die Koordination der ÜGK zu überdenken, und hierbei
   - eine eindeutige Trennung zwischen Auftraggeber *in und Projektleitung ÜGK zu schaffen
   - einen regelmässigen, formalisierten und transparenten Dialog zwischen den beiden Gremien zu implementieren
4. Die strategisch/politische ÜGK-Steuerung um eine wissenschaftliche Steuerung zu ergänzen, und hierbei
   - das geplante wissenschaftliche Konsortium (WiKo) zu implementieren
   - dieses Konsortium um einen wissenschaftlichen Beirat mit unabhängigen (inter)nationalen Expert*innen zu ergänzen.

Die Stellungnahme der EDK (vom 21.05.2019) zur Umsetzung der Empfehlungen in Massnahmen der Qualitätsentwicklung scheint die im Audit-Report benannten Empfehlungen nur in Teilen zu adressieren. So werden für die geplanten Aufgabenentwicklungen anscheinend keine zusätzlichen finanziellen Ressourcen zur Verfügung stehen. Stattdessen wird auf erhoffte oder vermutete Effizienzsteigerungen durch verbesserte Koordination, Erfahrungswerte und geringere Itemzielzahl verwiesen (EDK 2019, S. 23). Nach wie vor scheint nicht vorgesehen zu sein, Aufgabenentwicklung, Testzusammenstellung, Testdesign, Stichprobendesign, Standard-Setting und Datenanalyse *institutionell* stärker miteinander zu verzahnen, wie dies in Large-Scale-Assessment-Programmen anderer Länder (z.B. Luxemburg, Österreich oder Deutschland) der Fall ist. Hier liegen bisher eher unkonkrete Absichtsäusserungen seitens der EDK vor.

Ein Aufgreifen der Empfehlungen des Audits von Fischbach und Ugen (2018) im Bereich der Pilotierungen scheint hingegen besser realisiert worden zu sein. So sollen in künftigen ÜGK-Erhebungen grundsätzlich Pilotierungen, im Falle der erstmals administrierten ÜGK-4 auch Prä-Pilotierungen stattfinden (vgl. https://uegk-schweiz.ch/uegk-2024-harmos-4/). Hierbei wären aus Sicht des Gutachtenteams Budgetierungsfragen konkreter, frühzeitiger und verbindlicher zu klären, als dies derzeit von der EDK vorgeschlagen wird. Dies gilt ebenso für eine eventuelle Ausweitung auf naturwissenschaftliche Testdomänen oder die Testung produktiver Kompetenzen, die nach Aussage der Gesprächsteilnehmenden mit den bisherigen Mitteln nicht zu bewältigen seien. Grundsätzlich wird seitens der ADB der Wunsch geäussert, die technischen und operativen Aspekte der Skalierung extern zu vergeben, da andernfalls die Kapazitäten in der ADB nicht mehr ausreichten.

Die massiven sprachregionalen Unterschiede im Funktionieren einzelner Testaufgaben (Differential Item Functioning, „Sprachen-DIF“), die im Audit-Report von Fischbach und Ugen (2018) problematisiert werden und für die eine rigorose Qualitätssicherung des Übersetzungsprozesses empfohlen wird, werden seitens der EDK mit dem Verweis auf verbesserte iterative Checks innerhalb der Entwicklungsgruppe und stärkere Ko-Konstruktion der sprachregionalen Vertreter*innen der Entwicklungsgruppe adressiert. Unklar bleibt jedoch, ob formale Standards der Übersetzung von Testaufgaben berücksichtigt wurden, wie sie etwa in den Test Adaptation Guidelines (TAG) der International Test Commission (ITC 2005 und Hambleton & Zenisky 2010) oder im TRAPD-Ansatz (Translation, Review, Adjudication, Pretesting, and Documentation Approach) dargelegt werden. Darüber hinaus sollte bedacht werden, dass in einer mehrsprachigen Umgebung allfällige Unterschiede zwischen den Sprachen nicht nur als „Fehler“ interpretiert werden sollten, sondern wertvolle Hinweise auf unterschiedliche Kulturen geben können.

Es besteht ein Grundproblem in der Sequenz beziehungsweise im Timing des Testentwicklungsprozesses.

Insgesamt liegt trotz der genannten Anpassungen aus Sicht des Gutachtenteams ein Grundproblem in der Sequenz beziehungsweise im Timing des Testentwicklungsprozesses: Selbst wenn zukünftig vor jeder ÜGK-Hauptuntersuchung eine Pilotierung des Aufgabenpools an einer repräsentativen, grossen

Stichprobe stattfindet, können dennoch psychometrisch problematische Items nach eventuellen Korrekturen oder nachentwickelte Items nicht mehr vor dem Haupttest auf ihr „Funktionieren“ überprüft werden. In anderen Ländern (z.B. in Deutschland, siehe Pant & Stanat, 2015) ist hingegen zwischen Pilotierung (Fokus auf Qualität der Einzelaufgabe) und Haupterhebung eine Normierungsstudie (Fokus auf Eichung des gesamten pilotierten Aufgabenpools) geschaltet. Das Festlegen der Schwellenwerte für das Erreichen der Grundkompetenzen erfolgt so im Schweizer Konzept zwangsläufig anhand der Daten der Haupterhebung, während dies in Deutschland nach der Normierung passiert. Dies führt möglicherweise im Schweizer Design dazu, dass das Standard-Setting besonders anfällig dafür ist, „überraschende“ Verteilungen der Zielindikatoren (d.h. Anteil der Personen, die die Grundkompetenzen erreichen) quasi hinnehmen und erklären zu müssen oder politisch beeinflussen zu wollen.
Die alternativen Empfehlungen des Audits von Fischbach und Ugen (2018), entweder (a) die Standards selbst zu evaluieren, auf ihre (prognostische) Validität zu prüfen und allenfalls zu revidieren („Empirischer Ansatz“), (b) die Mindeststandards zu Regelstandards zu erklären („Terminologischer Ansatz“) oder (c) scheinbar zu „strenge“ Schwellenwerte stehen zu lassen und die resultierenden Verteilungen als Folge (noch nicht) umgesetzter Lehrplan-Implementation zu erklären („Rhetorischer Ansatz“) sind alle mit erheblichen Konsequenzen und Problemen behaftet.
Der – aus wissenschaftlicher Sicht zu favorisierende – „Empirische Ansatz“ setzt langwierige und teure Längsschnittstudien zur Beantwortung der Frage voraus, ob das (Nicht-)Erreichen der Grundkompetenzen tatsächlich zu differenziellen Bildungs- oder gar Erwerbsverläufen führt. Der „Terminologische Ansatz“ zieht die ebenfalls hochkomplexen Fragen der Entwicklung eines differenzierteren Standard- und Kompetenzstufenmodells nach sich und der „Rhetorische Ansatz“ ist ein möglicherweise ungedeckter Scheck auf die zukünftige Wirksamkeit gerade begonnener Curriculums-Umstellungen.
Gerade der „Rhetorische Ansatz” verweist auch auf eine Herausforderung, die weder im Audit von Fischbach und Ugen (2018) noch in den Gesprächen mit den Akteur*innen bisher ausreichend thematisiert erscheint: Wie soll für zukünftige Trendbetrachtungen (d.h. Vergleich des Anteils der grundkompetenten Schüler*innen im gleichen Fach und in der gleichen Jahrgangsstufe über die Zeit) im Bereich der Testentwicklung verfahren werden? Bisher wurden bis auf Anker-/Linkitems offenbar alle Testaufgaben veröffentlicht und sind somit für Folgeerhebungen „verbrannt“. Dies hätte für Trendbetrachtungen zur Folge, dass jeweils ein grösserer Teil des Aufgabenpools neu entwickelt werden muss. Bei den von Fischbach und Ugen (2018) aufgezeigten psychometrischen Problemen wäre eine inhaltliche Stabilität/Volatilität des Konstrukts und der Cut-Scores über die Erhebungswellen hinweg ein kritischer Validitätsaspekt, um Veränderungen über die Zeit sinnvoll interpretieren zu können.

Stärkung der Testentwicklung durch die Bildung multidisziplinärer Teams, durch eine stärkere fachdidaktische Fundierung oder Vermittlung von psychometrischen Grundkompetenzen gewünscht.

Moniert werden seitens der Gesprächsteilnehmer*innen aus der Testentwicklung und Umsetzung der ÜGK im Schulfeld, dass es nach wie vor keine nachhaltigen und personell ausreichend ausgestatteten sowie institutionell abgesicherten Strukturen der Item-/Testentwicklung gebe. Alle zeitkritischen Arbeiten seien damit in gewisser Weise weiterhin prekär angelegt und könnten sich im Falle einer Ausweitung des Domänenspektrums und/oder der Testfrequenz als dysfunktional erweisen. Die EDK (2019) hat als Reaktion auf die Empfehlungen des Audits von Fischbach und Ugen (2018) für die kommenden Erhebungen zwar den Anteil von Lehrpersonen in der Aufgabenentwicklung nach eigenen Angaben erhöht (vgl. EDK 2019, S. 23). Eine detaillierte Aufstellung dazu, wie viele Lehrpersonen in welchen Fächern mit welcher Expertise im Entwicklungsprozess beteiligt beziehungswiese eingeplant sind, lag dem Gutachtenteam nicht vor.
Zudem wird es in den durchgeführten Gesprächen seitens der ADB grundsätzlich als positiv bewertet, dass die relevanten Akteur*innengruppen aus verschiedenen Kantonen oder Sprachregionen an der

Aufgabenentwicklung beteiligt sind. Dennoch wird in den Gesprächen insbesondere die weitere Stärkung der Testentwicklung durch die Bildung echter multidisziplinärer Teams für die Testentwicklung (z.B. Pädagog*innen, Edumetriker*innen, Expert*innen für digitale Entwicklung) gewünscht. Andere Äusserungen gehen dahin, eine stärkere fachdidaktische Fundierung der Aufgabenformate vorzunehmen („Fachdidaktisch breit abgestützte Aufgabenentwicklung, die aktuellem fachdidaktischem Forschungsstand entspricht und nicht einfach der Erfassungsmethode geschuldet ist“).
Allerdings wird auch angeregt, dass den Beteiligten an der Aufgabenentwicklung mehr psychometrische Grundkenntnisse (IRT, Schwellenwerte) vermittelt werden. Ein universitärer Zertifikatslehrgang („CAS Kompetenzdiagnostik“) an den Universitäten Bern und St. Gallen kam bereits kurz nach dessen Einführung aufgrund mangelnden Interesses nicht mehr zustande, so dass eine Akquisition und systematische Grundausbildung neuer Aufgabenentwickler*innen auf diesem Wege bisher nicht erfolgreich war.

4.4 Technische Fragen zur ÜGK (B3.14; B3.17; B3.18)

Prozess des Standard-Setting ist State of the Art, teilweise bleiben aber Unklarheiten.

Das Audit von Fischbach und Ugen (2018) kommt explizit zum Ergebnis, dass der Prozess des Standard-Settings nach State of the Art erfolgt sei. Sowohl der gewählte Ansatz (modifizierte Bookmark-Methode) als auch die Panel-Zusammensetzung wie auch die Abläufe und die gewählten Response Probabilities (62.5 %) sind international üblich[17]. Mithin kann die teilweise kritische Rezeption der darauf basierenden Anteile von Getesteten, die die Grundkompetenzen (nicht) erreicht haben, nicht dem Verfahren oder seiner Implementation zugerechnet werden. In den Gesprächen wurden ebenfalls keine grundlegenden Zweifel an der Validität des Standard-Settings geäussert oder alternative Ansätze (z.B. Angoff-Verfahren) präferiert.
Es bliebe allerdings zu klären, ob die Panelteilnehmer*innen des Standard-Setting-Prozesses korrekt und über die verschiedenen Testdomänen gesehen einheitlich instruiert wurden (vgl. Fussnote 8 in Kapitel 3.2).

Ob auf ein digitales Assessment, durchgeführt von den Lehrpersonen, umgestellt werden soll, wird kontrovers gesehen.

Die Frage, ob die ÜGK mittelfristig auf ein digitales Assessment umgestellt werden sollte, *das von den Lehrpersonen selbst durchgeführt wird*, teilt sowohl Personen aus der Forschung und Aus- und Weiterbildung als auch kantonale Vertreter*innen in der standardisierten Befragung gleichermassen in zwei Gruppen: 46 % der kantonalen Vertreter*innen sind (eher) dafür, 36 % (eher) dagegen, bei Personen aus Forschung und Aus- und Weiterbildung fällt die Tendenz ähnlich, wenn auch in anderer Richtung aus, das heisst 36 % sind (eher) dafür beziehungsweise 40 % (eher) dagegen.
Zur Möglichkeit der kosteneffizienteren Testung mit schuleigenem Equipment äussern sich Vertreter*innen der ADB in den offenen Fragen grundsätzlich positiv, ohne dass dies jedoch konkreter ausgeführt würde („Weniger Geld ausgeben. Mit Testinfrastruktur der Schulen machbar.“).
Grundsätzlich ist zu bedenken, dass die Veränderung des Testmodus – von Testleitungs-administriert zu Lehrpersonen-administriert – zu erheblichen sogenannten „Mode-Effekten“ führen kann, die nur schwer zu kommunizieren sind. Die Erfahrungen aus dem Kontext der Vergleichsarbeiten (VERA)[18] in Deutschland zeigen, dass es für einzelne Aufgabentypen zu starken Schwierigkeitsschwankungen in

[17] Inhaltlich ist eine Response Probability von 62,5 % so zu interpretieren, dass die Teilnehmenden der Standard-Setting-Panels diejenige Testaufgabe in einem „Buch“ aller Testaufgaben, die nach aufsteigender empirischer Itemschwierigkeit geordnet wurden, identifizieren sollen, die ein/e „gerade so grundkompetente/r Schüler*in“ mit hinlänglicher Sicherheit (hier 62,5 %) richtig beantworten kann.

[18] https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsmonitoring/verfahren-zur-qualitaetssicherung-auf-schulebene/vera-faq.html [31. Oktober 2021]

Abhängigkeit davon kommen kann, ob eine Lehrperson oder eine externe Testleitung die Aufgaben administriert. Daher ist insbesondere darauf zu achten, dass in Pilotierung und Haupterhebung *der gleiche Modus* eingesetzt wird, um Inkonsistenzen zu vermeiden.

Verbesserte Dokumentation der Daten und der technischen Vorgehensweise gewünscht, sprachregionale Unterschiede in den Technischen Berichten und Dokumentationen sichtbar machen.

In den Antworten auf die (halb-)offenen Fragen der standardisierten Befragung finden sich mehrere Äusserungen von Forschenden, die sich eine verbesserte *Dokumentation der Daten* und der technischen Vorgehensweise wünschen, um Daten gegebenenfalls leichter für eigene Forschungsarbeiten nutzen zu können. Hervorgehoben wird explizit, dass zwecks besserer Verständlichkeit der Datenstrukturen die Dokumentation verbessert werden sollte, um zu einer externen Datennutzung zu motivieren („Datenformat, das nicht nur von Forscher*innen genutzt werden kann, die das Datenverarbeitungsmodell entwickelt haben [um zu vermeiden, dass die Daten nur von einigen wenigen Expert*innen genutzt werden]. Verschiedene Nutzungsebenen der Daten berücksichtigen“).
Insgesamt fällt auf, dass die Technischen Berichte und Dokumentationen[19] über die Sprachregionen, Testdomänen und Entwicklungsetappen hinweg teilweise nicht einheitlich gestaltet sind, was die vergleichende Rezeption und Beurteilung erschwert.

[19] Wir beziehen uns auf folgende Dokumente: Angelone & Keller (2019b), Erzinger et al. (2019), Pham et al. (2019a), Pham et al. (2019b), Salvisberg (2019), Verner & Helbling (2019).

5 Institutionelle Gestaltung der ÜGK

5.1 Organisation und Verantwortungsstruktur (B2.7; B.2.8; B2.9)

Trotz deutlicher Verbesserung nach wie vor verschiedene Probleme in der Projektorganisation sichtbar.

Aus den Gesprächen lassen sich als Ausgangspunkt in Bezug auf die Organisations- und Verantwortungsstruktur zwei Aussagen identifizieren, die mehrfach und aus unterschiedlichen Positionen eingebracht wurden: (1) Die Anfangsphase der ÜGK wurde als wenig koordiniert wahrgenommen. (2) Durch die Einrichtung des ICER, die übergreifende Koordination durch das ICER sowie die personelle Besetzung im Generalsekretariat der EDK hat sich die Situation deutlich gebessert. Diese Verbesserung wurde von mehreren Seiten wahrgenommen und in vielen Gesprächen explizit angesprochen. Nach wie vor bestehen allerdings Probleme hinsichtlich der Projektorganisation.

Es fehlt eine klare Abgrenzung der Kompetenzen zwischen dem Generalsekretariat der EDK und ICER.

Wie ein*e Gesprächsteilnehmer*in konkret formulierte, seien Abgrenzungen zwischen den Kompetenzen des GS EDK und ICER notwendig. Dieses Faktum wird auf unterschiedliche Weise immer wieder sichtbar, zum Beispiel in der Frage danach, wer denn eigentlich die ÜGK steuere. In anderer Form wurde dies in den Gesprächen evident, wenn das Fehlen einer „Identifikationsfigur“, die die ÜGK in der Öffentlichkeit repräsentiere, angesprochen wurde.
Es braucht klare Kompetenzdefinitionen für das GS EDK und das ICER innerhalb der ÜGK, so dass eine zentrale Steuerung des Gesamtvorhabens durch das GS EDK erhalten bleibt bei Gewährung der notwendigen Unabhängigkeit des ICER in der operativen und wissenschaftlichen Umsetzung. Dafür scheint es unumgänglich, die Ziele der ÜGK von Seiten der Politik klar zu definieren und unter Berücksichtigung des intendierten Wirkungsmodells einen langfristigen, verbindlichen Erhebungsplan auszuhandeln sowie mit wesentlichen Eckpunkten (etwa der Festlegung von Erhebungszeitpunkten, -zielgruppen und -inhalten) zu versehen. In diesem Rahmen benötigt das ICER eine entsprechend umfangreiche Unabhängigkeit für eine wissenschaftlich abgesicherte Implementation und Umsetzung.

Das ICER beschreibt seine Rolle als koordinierend. Die derzeitigen Strukturen führen zu einer Diffusion in der Leitung.

Das ICER beschreibt seine Rolle im Rahmen der ÜGK weder als Projektsteuerung noch als Projektleitung, sondern als koordinierend. Entscheidungen fallen gemäss den involvierten Akteur*innen entweder in der Projektleitung ÜGK als konsensuale Entscheide oder in der jeweils für einen Teilprozess zuständigen Arbeitsgruppe. Wer als Letztverantwortliche*r Entscheidungen trifft, wenn in der Projektleitung ÜGK kein Konsens erzielt werden kann, blieb in den Gesprächen offen. Damit ist die ÜGK durch eine sehr partizipative Entscheidungsstruktur geprägt. Dies mag den Vorteil von einem vermutlich teils intensiven entscheidungsvorbereitenden Diskurs haben, wobei anzumerken ist, dass ein entsprechend breiter Diskurs auch bei klaren Entscheidungsverantwortungen möglich ist. Jedenfalls erschwert diese Vorgehensweise die Koppelung von Entscheidungs- und Ergebnisverantwortung, den Umgang mit kurzfristig notwendigen Entscheidungen und einem eventuell erforderlichen Krisenmanagement. Ausserdem erschwert dies die Abstimmung von verschiedenen Teilprozessen und behindert effiziente, weitsichtige Entscheidungsprozesse. Koordination statt Projektsteuerung ist bei einem Unterfangen dieser Grösse und Komplexität als langfristig dysfunktional einzustufen. Durch diese Logik gibt es keine Möglichkeit, sichtbar zu werden, was wiederum auf das Problem der fehlenden Identifikationsfigur – die durchaus auch institutioneller Natur sein kann – verweist. Daher ist die Projektleitungsstruktur der ÜGK insbesondere unter dem Gesichtspunkt klarer Entscheidungs- und Ergebnisverantwortung zu diskutieren, zumal in der Kooperationsvereinbarung zwischen der EDK und der Universität

Bern (EDK, 2018) explizit steht, dass die Universität Bern beziehungsweise das ICER die *Gesamtleitung der Prozesse* innehabe (Absatz 3.2.2 in der Kooperationsvereinbarung). Im Gegensatz dazu stehen die Formulierungen zu den definierten Aufgaben des wissenschaftlichen Konsortiums (Art. 7, Abs. 2) im ÜGK-Reglement (EDK, 2014), auf welche im relevanten Abschnitt (Absatz 3.1 in der Kooperationsvereinbarung) einleitend verwiesen wird. Hier wird vorrangig von *Koordination* gesprochen. Diese Ambivalenz zwischen Leitung und Koordination ist demnach bereits in den Vertragsdokumenten grundgelegt.

Bestehende Reglemente sind mit den Organisationsstrukturen nicht immer im Einklang, insbesondere bezüglich des Wissenschaftlichen Konsortiums und der Entwicklung des Kontextfragebogens.

Besonders diffus erscheint die Position des wissenschaftlichen Konsortiums. Es existiert als zusätzliches Gremium und wird als loser Verbund beschrieben, der vor allem zur breiteren Information genutzt wird. Damit ist das wissenschaftliche Konsortium nicht das strategische, wissenschaftliche Gremium wie etwa im ÜGK-Reglement beschrieben und seine konkrete Rolle ist unklar. Die Aufgaben, die im ÜGK-Reglement einem wissenschaftlichen Konsortium zugeschrieben wurden, sind zwischenzeitlich im Rahmen der Kooperationsvereinbarung zwischen dem GS EDK und dem ICER Letzterem zugewiesen worden. Die Organisationsstrukturen für die ÜGK haben sich in den letzten Jahren weiterentwickelt, so dass sich der Bedarf ergibt, die (immer noch geltenden) Reglemente aus der Anfangszeit der ÜGK mit den zwischenzeitlich gewachsenen Strukturen in Einklang zu bringen. Dies dient der Klarheit in den Verantwortungsstrukturen und der Verwendung von Begriffen.
Die unklaren Leitungs- und Entscheidungsstrukturen sowie die unscharfe Abgrenzung zwischen Politik und Wissenschaft werden am Beispiel des Kontextfragebogens besonders deutlich. Der Ablauf zur Entwicklung des Kontextfragebogens – abgesehen von den für die Skalierung relevanten Basic-Variablen zum sozio-ökonomischen Hintergrund der Schüler*innen – wird in den Gesprächen als sehr offener Prozess beschrieben: Interessierte Institutionen könnten Wünsche und Ideen einbringen, einige funktionierende Skalen würden von vorausgegangenen ÜGK-Erhebungen übernommen. Dahinter stehe die Idee, dass interessierte Wissenschaftler*innen Skalen einbringen und diese dann auch selbst für Auswertungen nutzen können. Letzteres werde durch die Komplexität des resultierenden Datensatzes sowie die Einschränkungen im Umfang des Kontextfragebogens jedoch behindert. Insgesamt wirkt bei der Entwicklung des Kontextfragebogens die Organisations- und Entscheidungsstruktur besonders unklar.
Das ist zum einen im Zusammenhang mit den nicht genügend abgestimmten Vorgaben und Interpretationen zu sehen, wozu die Kontextfragebögen im Kontext der ÜGK dienen sollen. Den Ausgangspunkt zur Entwicklung des Kontextfragebogens bildet nicht ein gemeinsames Verständnis über den Zweck des Kontextfragebogens, sondern ein Konglomerat von Ansprüchen an den Kontextfragebogen, so beispielsweise von den Expert*innen zur Skalierung der Daten (wofür zentrale Variablen des Kontextfragebogens für das Hintergrundmodell benötigt werden) oder von Wissenschaftler*innen, die mittels des Kontextfragebogens eigene Forschungsinteressen realisieren möchten. Eine Perspektive, die in den Gesprächen nicht genannt wurde, aber eigentlich zentral erscheint, sind Bedürfnisse in Bezug auf die Ergebnisberichterstattung zur ÜGK.
Zum anderen werden in den Prozessen rund um den Kontextfragebogen jedoch auch die insgesamt diffusen Entscheidungsprozesse sichtbar. Wer trägt die Verantwortung dafür, dass die im Rahmen des Kontextfragebogens erhobenen Informationen die Bedürfnisse der ÜGK erfüllen? Die Tatsache, dass die Aufnahme einzelner Items in den Kontextfragebogen letztlich durch den EDK-Vorstand genehmigt wird, kann als ein Hinweis für die Vermengung von Verantwortung zwischen Politik und Wissenschaft gesehen werden, was dazu führt, dass die Politik – statt weitsichtiger strategischer Entscheidungen – operative Einzelentscheidungen trifft.

Die Trennung zwischen ICER und ADB und damit auch die Trennung der Verantwortung für Aufgabenentwicklung einerseits und Durchführung der ÜGK andererseits wird kritisiert.

Die Trennung zwischen den Verantwortlichkeiten für die Test- und Aufgabenentwicklung auf der einen Seite und für die Durchführung der Erhebungen auf der anderen Seite wird übereinstimmend von allen an den Gesprächen beteiligten Akteur*innen auf der operativen Ebene als problematisch gesehen. Dies ist insofern kritisch einzuschätzen, als die Aufgabenentwicklung das konzeptionelle Kernelement der ÜGK darstellt. Durch die institutionelle Verantwortungstrennung liegt dieser Kernprozess nicht in der Steuerungsverantwortung der mit der Durchführung beauftragen Institution. Entscheidungen in Bezug auf die Aufgaben- und Testentwicklung müssen sehr früh im Vorbereitungs- und Durchführungsprozess der ÜGK getroffen werden. Sie stehen mit verschiedenen (chronologisch später angesiedelten) Teilprozessen in Wechselwirkung. Das bedeutet, dass Entscheidungen im Kontext der Aufgabenentwicklung Auswirkungen auf die Prozesse der Erhebung, Verarbeitung und Analyse der Daten haben. Spezifische Anforderungen dieser später angesiedelten Prozesse müssen in Entscheidungen zur Test- und Aufgabenentwicklung Berücksichtigung finden. Die Verantwortungstrennung zwischen diesen und dem Aufgaben- und Testentwicklungsprozess beeinträchtigt daher sowohl Effektivität und Effizienz als auch die Qualität der Produkte.

Die ADB ist innerhalb der ÜGK isoliert, eine institutionelle Anbindung an eine wissenschaftliche Institution fehlt

Die ADB in der derzeitigen Verortung wird sowohl aus einer ÜGK-Innen- als auch aus einer -Aussenperspektive als isoliert beschrieben. Das betrifft sowohl eine strukturelle Isolation innerhalb der Projektorganisation der ÜGK als auch einen Mangel an institutioneller wissenschaftlicher Anbindung. Gleichzeitig ist die ADB mit der Aufgaben- und Testentwicklung für den Kernbereich der ÜGK zuständig, weil mit der Aufgabenentwicklung wesentliche inhaltliche Festlegungen erfolgen. Das verträgt sich weder mit ihrer isolierten Position noch mit der engen Anbindung an das GS EDK. Während das Festlegen von Anforderungen an das Bildungssystem (in Form von Bildungsstandards) in der Verantwortung der Politik liegt, betrifft das in der ADB angesiedelte Aufgabenportfolio die wissenschaftsbasierte Operationalisierung dieser Anforderungen. Sowohl die Entwicklung von Tests und Aufgaben als auch die Setzung von Schwellenwerten, die das Erreichen des inhaltlich festgelegten Anforderungsniveaus kennzeichnen, sind Aufgaben, die nach wissenschaftlichen Kriterien zu bearbeiten sind. Insbesondere vor dem Hintergrund von Glaubwürdigkeitsfragen ist eine formale Trennung zwischen den politischen Entscheiden (was sind die Grundkompetenzen, die Schüler*innen jedenfalls erreichen sollten?) und der wissenschaftlichen Operationalisierung dieser Entscheide (der operativen Umsetzung für die Erhebung des Erreichens der Grundkompetenzen) wichtig.
Darüber hinaus ist die ADB personell unterbesetzt und bietet qualifiziertem Personal wenig Perspektiven. Wesentliche Informationen zum Kernprozess der Test- und Aufgabenentwicklung liegen bei der ADB. Durch ihre vom Rest der ÜGK-Durchführung abgekoppelte Positionierung führt dies zu einer prekären Situation. Das Know-how liegt bei einigen wenigen Personen (derzeit bei einer Person!). Bei personellen Wechseln gibt es keine Strukturen, die den persönlichen Know-how-Verlust durch institutionelles Know-how abfedern können. Die isolierte Position der ADB (verstärkt durch deren geringe personelle Besetzung) birgt daher ein grosses Risiko in Bezug auf kritischen Know-how-Verlust.
Aus diesen Gründen sollte eine stärkere institutionelle Integration der ADB in die ÜGK überlegt werden. Relevant erscheinen insbesondere eine institutionelle Anbindung an eine wissenschaftliche Institution und eine Unabhängigkeit vom GS EDK. Letzteres auch, weil – wie oben angesprochen – die Test- und Aufgabenentwicklung die Operationalisierung der Festlegungen, die politisch zu treffen sind, darstellt. Ähnlich verhält es sich bei der Schwellenwertsetzung, die auf Basis einer inhaltlichen Definition der Grundkompetenzen (politische Entscheidung) ein operativer Prozess ist, der nicht politisch beeinflusst sein sollte. Jedenfalls bedarf es einer besseren institutionellen Anbindung an die wissenschaftliche Community, vorzugsweise verbunden mit einer besseren Integration in den Gesamtprozess der ÜGK.

Die Entscheidungen im Bereich der Aufgabenentwicklung sollten von Entscheidungen der Durchführung der ÜGK nicht entkoppelt sein. Es gilt umfangreiche Wechselwirkungen zwischen Konzeptions- und Durchführungsaspekten zu berücksichtigen, was durch die Trennung der Verantwortlichkeiten massiv erschwert wird. Dies könnte durch eine institutionelle Integration der ADB in das ICER (als für die Projektsteuerung verantwortliche Institution) unterstützt werden.

Die fehlende zeitliche Perspektive verhindert eine langfristige Planung und konzeptionelle Gestaltung der ÜGK.

Der Bedarf an einem langfristigen verbindlichen Plan für die ÜGK-Erhebungen wird in verschiedenen Kontexten angesprochen und ist wohl die in den Gesprächen am häufigsten genannte notwendige Veränderung für die Zukunft. Die fehlende langfristige Perspektive steht dem Aufbau einer effizienten Organisationsstruktur im Wege. Für eine koordinierte Abwicklung und effiziente Konzeption der Erhebungen ist es erforderlich, frühzeitig über verbindliche Informationen zu wesentlichen Eckpunkten (langfristige Festlegung des Erhebungsdesigns, insbesondere Zeitpunkte, Schulstufen, Fächer und deren Abfolge) zu verfügen. Sowohl der Effizienz als auch der Nachhaltigkeit sind durch die derzeitige Situation mit sehr kurzfristigen Entscheidungen, kurzen Vorlaufzeiten für einzelne Erhebungen sowie dem Fehlen eines längerfristigen Plans deutliche Grenzen gesetzt. Dies betrifft auch die vertragliche und damit finanzielle Situation. Eine längerfristige Planung sowie eine damit zusammenhängende vertragliche Beauftragung der Institutionen dienen der Planungssicherheit. Durch die kurzen Vertragslaufzeiten besteht in allen Bereichen die Gefahr, kompetente Personen zu verlieren. Dies führt – nicht nur in Bezug auf die ADB – zur permanenten Gefahr von Know-how-Verlust und behindert den Aufbau eines stabilen Kompetenzzentrums. Die Notwendigkeit eines langfristigen, verbindlichen Plans gilt nicht nur für die direkt in die Vorbereitung und Durchführung der ÜGK involvierten Institutionen, sondern auch für die Verantwortlichen der Bildungsplanung in den Kantonen oder für Wissenschaftler*innen, die Forschung im Kontext oder unter Einbeziehung der ÜGK planen.

Die Einbindung der Sprachregionen ist erst in Teilen realisiert.

Vor allem von französischsprachigen Personen wurde in den Gesprächen auf die Wahrnehmung mangelnder Einbindung aller Sprachregionen hingewiesen. Die ÜGK wird von manchen als „deutschsprachiges“ Projekt wahrgenommen, zu dem aus den französischsprachigen Regionen schwer Zugang gefunden werden könne, beziehungsweise das Personal aufgrund von Sprachkompetenzen anstatt Expertise ausgewählt werden müsse, um mitreden zu können.

Die systematische Einbindung der Fachdidaktiken ist auf die Aufgabenentwicklung beschränkt.

Vertretungen der Fachdidaktiken sind ausschliesslich in den Bereich der Aufgabenentwicklung systematisch eingebunden. In dieser Rolle sind sie über die ADB an die Projektleitung ÜGK angebunden. Die ADB ist eher technisch-psychometrisch ausgerichtet, dies sowohl operativ als auch konzeptionell in Bezug auf das Aufgabenportfolio. Dies ist wichtig für die Erfüllung ihrer Aufgaben im Bereich der Aufgaben- und Testentwicklung sowie Skalierung und Schwellenwertsetzung. Allerdings führt dies dazu, dass die Perspektive der Fachdidaktiken derzeit vorrangig in die Aufgabenentwicklung einfliesst, nicht aber in andere Phasen der ÜGK, etwa in gesamtkonzeptionelle Aspekte oder die Ergebnisanalyse und -interpretation, da die Fachdidaktiken nur über die ADB in der Projektleitung ÜGK vertreten sind und damit nur indirekte Verbindungen mit anderen Teilprozessen der ÜGK-Erhebungen bestehen. Zwar scheint es verschiedene persönliche Kanäle zum GS EDK, zu Kosta HarmoS oder zur Projektleitung ÜGK zu geben. Eine verlässliche, transparente, institutionalisierte Basis für die vollumfängliche Berücksichtigung der wichtigen fachdidaktischen Perspektive in der ÜGK besteht allerdings nicht.

Deshalb sollten Strukturen gefunden werden, die die Berücksichtigung der Perspektive der Fachdidaktiken in der Gesamtkoordination der ÜGK ermöglichen. Dies sollte systematisch erfolgen, so dass die Perspektive der Fachdidaktiken in der Gesamtkoordination gezielt einbezogen wird und dies nicht abhängig von der Mehrfachverantwortung oder persönlichen Kanälen einzelner Personen ist. Dies könnte auch dazu beitragen, die derzeit als erziehungswissenschaftlich-soziologisch fokussiert wahrgenommene Daten- und Ergebnisnutzung um eine fachdidaktische Perspektive anzureichern. Damit wird insbesondere eine Brücke zum Unterricht und zur Lehrer*innenaus- und -weiterbildung geschlagen. Die breitere Einbeziehung fachdidaktischer Positionen in der ÜGK ist darüber hinaus wichtig, um der Segmentierung von Entscheidungsprozessen entgegenzuwirken. In dieser Hinsicht ist mit positiven Effekten in beide Richtungen zu rechnen – für die ÜGK als Ganzes und als Rückwirkung auf die Arbeit der Fachdidaktiken im Bereich der Aufgabenentwicklung. Dabei ist relevant, dass es unterschiedliche Fachdidaktiken zu berücksichtigen gilt und nicht „die Fachdidaktik“. Strukturelle Anbindungsmöglichkeiten sollten also auf die Unterschiedlichkeit etwa zwischen Mathematik und den Sprachen (nicht nur, aber auch vor dem Hintergrund eines mehrsprachigen Landes) Rücksicht nehmen. Darüber hinaus ist auch die derzeit fehlende Planungssicherheit bezüglich der ÜGK-Erhebungen ein hinderlicher Faktor für den langfristigen Aufbau von Expertise in Bezug auf die Test- und Aufgabenentwicklung in den Fachdidaktiken und eine systematische Einbindung der Fachdidaktiken in den Gesamtprozess der ÜGK.

5.2 Verhältnis ÜGK und Wissenschaft (A1.6; B2.10; B3.16; B3.19)

Das Verhältnis zwischen ÜGK als politisches Instrument der Systemevaluation und der Wissenschaft ist an vielen Stellen unklar.

Im Kontext der Frage zu den Beziehungen zwischen Wissenschaft und ÜGK stellt sich die Grundsatzfrage, inwieweit eine Ausrichtung der ÜGK auf die Wissenschaft erstrebenswert ist. Aus der Sicht von vielen Personen, die in den Gesprächen Kantone oder Durchführungsinstitutionen repräsentiert haben, erscheint die vorrangige Erwartung an die ÜGK eine Orientierung an den Bedürfnissen der Kantone beziehungsweise am Zweck des Monitorings zu sein. Um diesem Wunsch zu entsprechen, bedarf es einer vorrangigen Ausrichtung der Konzeption und des Designs der Erhebungen an den Bedürfnissen des Monitorings. Das betrifft sowohl die Konzeption der Tests als auch der Kontextfragebögen. Grundsätzlich ist eine wissenschafts-basierte Konzeption und Durchführung für eine erfolgreiche Umsetzung der ÜGK unbedingt erforderlich. Ob und inwieweit die ÜGK allerdings auf die Wissenschaft und die Bedürfnisse von Forschenden allgemein zugeschnitten sein soll, ist zu diskutieren.
Auf Basis der standardisierten Befragung werden im Folgenden die Sichtbarkeit der ÜGK, die Vernetzung des ICER innerhalb der Wissenschaft in der Schweiz und die Nutzung von ÜGK-Berichten und Daten durch Wissenschaftler*innen beschrieben[20].
Die Sichtbarkeit des ICER als Koordinationszentrum der ÜGK im Kontext der Bildungsforschung wird von den befragten Wissenschaftler*innen in der Summe nicht besonders hoch eingeschätzt. Deutlich mehr als die Hälfte der Befragten halten das ICER im Kontext der Bildungsforschung für überhaupt nicht oder eher nicht sichtbar (19 % überhaupt nicht sichtbar, 41 % eher nicht sichtbar). 16 % nehmen das ICER als eher sichtbar wahr. Als sehr sichtbar wird das ICER nur von einer Person eingeschätzt. Diese Ergebnisse sind jedoch vor dem Hintergrund der noch jungen Geschichte des ICER zu sehen.
Für ihre eigene Forschung ist der Austausch mit den wissenschaftlichen Institutionen der ÜGK (ICER, ADB) für etwas weniger als die Hälfte der befragten Personen eher oder sehr wichtig (23 % beziehungsweise 20 %). 14% halten diesen Austausch für ihre eigene Forschung für überhaupt nicht wichtig, fast

[20] Es sei darauf verwiesen, dass hier nur die Ergebnisse der Gruppe der 64 Forschenden (ohne Personen von der Aus- und Weiterbildung) berichtet werden. Weiter hat ein Anteil von circa 23 % der Forschenden diese Fragen nicht beantwortet.

ein Fünftel für eher nicht wichtig. Dabei sind deutliche Unterschiede zwischen den Disziplinen zu beobachten. Während 24 % der Befragten, die sich (auch) dem Bereich Erziehungswissenschaft zuordnen, den Austausch mit den Institutionen der ÜGK als sehr wichtig bezeichnen, sind dies unter den Fachdidaktiker*innen nur 14 % und niemand von jenen Personen, die ausschliesslich andere Disziplinen angegeben haben.
Dieser Austausch soll aus Sicht der Befragten zu je etwa gleichen Teilen (eher) verstärkt werden oder wird als gerade richtig eingestuft. Nur eine Person möchte den Austausch (eher) reduzieren. Dabei zeigt sich erwartungskonform, dass jene Personen, die den Austausch bisher als wichtig angesehen haben, diesen auch eher verstärkt wissen möchten.
Etwa ein Drittel der Wissenschaftler*innen äussert im Rahmen der standardisierten Befragung Interesse an einer Mitarbeit. Etwas grösser ist das Interesse, sich an der Konzipierung des Kontextfragebogens zu beteiligen (39 % eher oder sehr interessiert) als an der Entwicklung der Leistungstests (30 % eher oder sehr interessiert). Interesse an einer Beteiligung an der Entwicklung von Leistungstests und/oder Kontextfragebögen wird verstärkt von Personen geäussert, die den Austausch mit den Institutionen der ÜGK bereits schätzen gelernt haben sowie Ergebnisse oder Daten der ÜGK bereits genutzt haben.

Der Zweck des Kontextfragebogens ist ungeklärt und die resultierenden Daten sind derzeit weder für Monitoring- noch Forschungszwecke gut nutzbar.

Wesentliche Designentscheidungen sollten vor dem Hintergrund des Hauptzwecks der ÜGK als Instrument des Bildungsmonitorings erfolgen. Konzeption und Durchführung der ÜGK-Erhebungen sollten auf diesen Zweck ausgerichtet werden. Dies hat vor allem Auswirkungen auf die Arbeiten in Bezug auf den Kontextfragebogen. Wenn dem Zweck des Monitorings entsprochen werden soll, sind die Inhalte des Kontextfragebogens auch vorrangig auf diese Bedürfnisse abzustimmen. Dann braucht es eine zentral verantwortliche (wissenschaftliche) Stelle, die den Kontextfragebogen unter diesen Gesichtspunkten konzipiert und optimiert. Das setzt jedoch voraus, die Leistungen der Wissenschaft als Dienstleistung zu organisieren und zu alimentieren und auch für diesen Bereich eine klare Verantwortungstrennung zwischen Politik und Wissenschaft vorzunehmen. Die Annahme, Wissenschafter*innen könnten Serviceleistungen aus wissenschaftlichem Interesse erbringen, ist nicht erfüllbar. Die derzeitige Vorgehensweise, Wissenschaftler*innen offen einzuladen und zu fragen, was sie interessiert, und damit den Inhalt des Kontextfragebogens in erster Linie unsystematisch vom Forschungsinteresse einzelner Personen abhängig zu machen, ist dem Zweck des Monitorings nicht dienlich. Diese Vorgehensweise scheint – so die Aussagen in den Gesprächen – dazu zu führen, dass die resultierenden Daten für niemanden wirklich gut nutzbar sind. Für Monitoringzwecke scheinen die Inhalte des Kontextfragebogens eher zufällig verwendbar; für wissenschaftliche Zwecke lassen sich die Ergebnisse nur begrenzt nutzen, weil – so eine an den Gesprächen beteiligte Person – durch den beschränkten Platz und ohne inhaltliche Fokussierung keine für hochrangige Publikationen brauchbaren Skalen resultierten.
Eine Fokussierung in Designfragen auf den Zweck des Monitorings schliesst jedoch nicht aus, dass die entstehende Datenbasis für eine breitere wissenschaftliche Nutzung von Interesse sein kann. Dies betrifft sowohl eine fachdidaktische Nutzung von Ergebnissen auf Itemebene, als auch breitere wissenschaftliche Fragestellungen, die von der soliden Messung von Schüler*innen-Leistungen profitieren können. Was die Attraktivität der resultierenden Daten für die Wissenschaft angeht, wird als Einschränkung auf die Messung des Erreichens des einen Schwellenwerts hingewiesen. Nachdem dieser 0/1-Indikator in Bezug auf das Erreichen der Grundkompetenzen mithilfe einer dahinterliegenden kontinuierlichen Leistungsskala erzeugt wird, stehen jedoch grundsätzlich sehr wohl Testscores zur Verfügung, die als Achievement-Daten nutzbar sind (wiewohl in den meisten Fächern nicht mit einer Abdekkung des gesamten Leistungsspektrums; vgl. Kapitel 3.2).

Eine Nutzung der ÜGK-Daten und Ergebnisse durch die Wissenschaft ist erst punktuell sichtbar.

Geht es um die Klärung der Frage nach der Nutzung von ÜGK-Berichten sowie Daten durch Wissenschaftler*innen, können Informationen aus zwei Quellen genutzt werden. Zum einen hat eine Nachfrage bei FORS ergeben, dass bis Juni 2021 für 56 Forschungsvorhaben die Nutzung der ÜGK-Daten beantragt wurde. Dabei sind 21 Anfragen aus dem ICER selbst oder aus Institutionen der Projektleitung ÜGK erfolgt. Zwei Anfragen erfolgten in Bezug auf die Erstellung des Bildungsberichts.
Zum anderen können wiederum die Antworten aus der standardisierten Befragung herangezogen werden. 17 % der befragten Wissenschaftler*innen haben die Berichterstattung zu den Ergebnissen der ÜGK bereits als Ausgangspunkt für eigene Forschungsvorhaben genutzt; weitere 5 % geben an, dies zu planen. Ähnlich klein ist die Gruppe an Personen mit 17 %, die angeben, FORS-Daten für eigene Analysen verwendet zu haben. Personen, die FORS-Daten bereits genutzt haben, ziehen das auch für die Zukunft wieder in Erwägung. Von jenen, die bisher von dieser Möglichkeit noch keinen Gebrauch gemacht haben, plant immerhin mehr als die Hälfte, das in Zukunft (eventuell) zu tun. Ähnlich verhält es sich auch bei Personen, die die Berichterstattung bereits als Ausgangspunkt für ihre Forschung verwendet haben. Auch sie schliessen die Nutzung von FORS-Daten in Zukunft nicht aus. Nur etwa ein Viertel der Befragten gibt an, eine Nutzung von FORS-Daten nicht zu planen. Das sind durchwegs Personen, die in der Vergangenheit weder die Berichterstattung zu den Ergebnissen noch die FORS-Daten im Rahmen ihrer Forschungstätigkeit genutzt haben.
Erwartungskonform gibt es deutliche Zusammenhänge zwischen den berichteten Merkmalen. So sind insbesondere Personen, die bereits mit FORS-Daten oder der Ergebnisberichterstattung gearbeitet haben, zur Mitarbeit in der Entwicklung der Leistungstests oder der Konzipierung der Fragebögen bereit. Sie beurteilen auch die Sichtbarkeit des ICER im Kontext der Bildungsforschung besser und schätzen den Austausch mit den Institutionen der ÜGK mehr.

Änderungen im Konzept oder in den Instrumenten der ÜGK könnten eine bessere Nutzung der ÜGK-Daten ermöglichen.

38 % der Forschenden können sich Änderungen im Konzept oder in den Instrumenten der ÜGK vorstellen, die dazu führen würden, dass sie die Daten der ÜGK häufiger für die eigene Forschung nutzen könnten und dass damit eine breitere Einbeziehung der Forschenden erreicht werden könnte. 39 % der Forschenden sind der Ansicht, dass es Änderungen geben müsste, damit mehr oder relevantere Fragen der Bildungsforschung bearbeitet werden könnten.
In den Antworten auf die (halb-)offenen Fragen des Fragebogens finden sich mehrere Äusserungen von Forschenden, die eine bessere Zugänglichkeit der Daten, eine Synopse, Gegenüberstellung beziehungsweise Verknüpfung der Daten und Ergebnisse der ÜGK mit weiteren Quellen (kantonalen Erhebungen, Schulstatistiken) sowie eine bessere Aufbereitung und Vergleichbarkeit der Berichte unter Transferaspekten wünschen, um die Datennutzung zu verbessern.
Die Gespräche haben verschiedene Punkte aufgezeigt, wo spezifische Herausforderungen bei der Datennutzung bestehen. So liegt erst seit Frühjahr 2018 ein Datennutzungskonzept vor. Zudem ist das Einholen von Bewilligungen schwerfällig und selbst das ICER musste für die Nutzung der Daten entsprechende Bewilligungen einholen. Darüber hinaus sei für sekundäranalytische Auswertungen nach Ansicht einiger Gesprächsteilnehmenden der Merkmalskranz, der im Kontextfragebogen erhoben wird, zu wenig ausdifferenziert (sowohl hinsichtlich der Zahl der Konstrukte als auch der Zahl der Items pro Konstrukt), um damit in wissenschaftlich angesehenen Journalen publizieren zu können. Dies betrifft auch die fachdidaktische Forschung.
Die Attraktivität der Datensätze für eine sekundäranalytische Nachnutzung, die nicht nur auf statistisch-methodische Fragestellungen beschränkt bleibt, steht und fällt mit dem Umfang, dem Design und der Qualität der erhobenen Kontextdaten. Die Beispiele aus anderen Ländern zeigen eindrücklich, dass die Erhebung über eine Lehrpersonen-Stichprobe, ein Klassenstichprobendesign zur Erfassung von

Lerngruppenmerkmalen und -effekten sowie ein Eltern- und Schulleitungsfragebogen nicht nur die Ergebnisse des Bildungsmonitorings in ihrer Aussagekraft deutlich ausweiten, sondern auch zu vielen erfolgreichen Forschungsanträgen und hochwertigen Publikationen führen.
Um theoriegeleitete Ergänzungsstudien an zukünftige ÜGK-Erhebungen anzudocken und Forschungsvorhaben damit ebenfalls für eine SNF-Förderung potenziell zu qualifizieren, bräuchte es nach Ansicht der Gesprächsteilnehmenden eine vertragliche Kooperation mit weiteren Hochschulen und deren Forschungsgruppen, um auch hochschulseitig ein Commitment zur besseren Sekundärnutzung der Daten zu erzeugen.

6 Empfehlungen

Die Frage nach der bestmöglichen Strategie für die zielorientierte Entwicklung eines Bildungssystems kann als eine der wesentlichen Fragen der letzten 50 Jahre bezeichnet werden. Dominierte in der Schweiz wie auch in anderen deutschsprachigen und europäischen Ländern lange die Strategie, den sogenannten „Input“, also die personellen oder finanziellen Ressourcen des Bildungswesens oder die Lehrpläne und Qualität der Ausbildung von Lehrpersonen zu fokussieren und gegebenenfalls anzupassen, so haben seit den 2000er-Jahren die sogenannten „Output“-Steuerungssysteme an Relevanz gewonnen (Klieme et al., 2003). Diese beinhalten die Festlegung von Bildungsstandards und die regelmässige Überprüfung der Kompetenzen der Schüler*innen und damit die Analyse der Frage, ob die Schüler*innen zu einem bestimmten Zeitpunkt auch tatsächlich die in den Bildungsstandards festgelegten Ziele erreicht haben.

Auch wenn in der Schweiz (oder in den anderen deutschsprachigen Ländern) kein radikaler Paradigmen-Wechsel von einer Input- zu einer Output-Steuerung stattgefunden hat (Altrichter & Maag Merki, 2016), sondern vielmehr die bereits früher existierenden Input-Steuerungsinstrumente durch Elemente der Leistungserreichungsüberprüfung (Output) erweitert wurden, so zeigen dennoch viele Studien, dass die neuen respektive erweiterten Monitoringsysteme einen – teils fundamentalen – Einfluss auf die Bildungspraxis, die Bildungsadministration und -politik haben können (Maag Merki, 2016). Sie haben das Potenzial, die Interaktionen zwischen den Akteur*innen sowie teilweise auch die Prozesse im Unterricht und in der Schule zu verändern. Darüber hinaus kann in vielen Ländern beobachtet werden, dass neue Institutionen entstehen oder Fachstellen implementiert werden, um ein stärker auf die Lernergebnisse ausgerichtetes Monitoringsystem überhaupt umsetzen zu können. Für die Schweiz sind dies auf nationaler Ebene am offensichtlichsten das ICER und die ADB, aber auch die Verantwortlichen für das Bildungsmonitoring in den Kantonen oder „kantonale Referenzpersonen“, die die Umsetzung der ÜGK in der Praxis unterstützen sollen. Diese müssen ihren eigenen Aufgaben- und Zuständigkeitsbereich im Verhältnis zu den neuen Aufgaben sowie zu bereits existierenden Personen und Institutionen entwickeln. Entscheidungszuständigkeiten und Aufgabenbereiche müssen aufeinander abgestimmt und neu ausgehandelt werden.

Ein Blick in die umfangreiche Forschungsliteratur im Zusammenhang mit der Effektivität der neuen oder angepassten Monitoringstrategie macht deutlich, dass die neue Form des Bildungsmonitorings sehr anspruchsvoll und voraussetzungsreich ist. Keinesfalls bestehen ‚einfache Wege‘, die die Entwicklung eines guten Bildungssystems, in dem (möglichst) alle Schüler*innen die gesetzten Bildungsstandards und Bildungsziele erreichen können, garantieren (Maag Merki, 2016). Diese Komplexität wird zudem dadurch gesteigert – und das kann im internationalen Vergleich, mit wenigen Ausnahmen, durchaus als ein Alleinstellungsmerkmal der Schweiz bezeichnet werden – dass das Erreichen der Grundkompetenzen durch die Schüler*innen nicht nur zu verschiedenen Jahrgangsstufen und in verschiedenen Fächern/Fachbereichen zu überprüfen ist, sondern dass dies auch sprachübergreifend zu geschehen hat.

Noch ein weiterer Aspekt kann als spezifisch schweizerisch betrachtet werden: Im Zuge der Umsetzung des Art. 62 der Bundesverfassung und dem Auftrag zur Harmonisierung von zentralen Eckwerten der obligatorischen Schule steht die Überprüfung des Erreichens der Grundkompetenzen im Kontext der Harmonisierung, auch wenn bislang wenig klar ist, inwiefern diese Harmonisierung über das Festlegen gemeinsamer Bildungsziele hinausgehen soll.

Für die Schweiz bedeuten die Einführung von Bildungsstandards und die regelmässige Überprüfung des Erreichens der Grundkompetenzen der Schüler*innen zudem eine zweifache Neuerung. Das erste Mal werden auf nationaler Ebene Bildungsstandards für unterschiedliche Jahrgangsstufen in unterschiedlichen Fächern/Fachgruppen festgelegt und empirisch überprüft. Zugleich wird mit der ÜGK das erste Mal überhaupt eine nur für die Schweiz entwickelte Bildungsmonitoring-Strategie auf nationaler Ebene implementiert. Dies in einem Land, in dem das Bildungswesen in der obligatorischen Schule in hohem Masse föderalistisch geprägt ist und bisherige Angleichungen – wenn überhaupt – in den Sprachregionen oder zwischen einzelnen Kantonen realisiert worden sind.

Die Analysen der durchgeführten Gespräche, die Auswertung der Rückmeldungen aus der standardisierten Befragung sowie die Analyse bestehender Dokumente und Reglemente ermöglichen es, einen differenzierten Blick in die Strukturen, Prozesse und Potenziale der aktuellen Umsetzung der ÜGK zu erhalten. Aus dem Zusammenführen dieser Analyse werden in der Folge Empfehlungen für die Weiterentwicklung der ÜGK formuliert.

Die folgenden Themenbereiche werden von der Gutachtengruppe als essenziell für eine zukunftsträchtige Umsetzung der ÜGK erachtet. Wir haben uns zum einen eng an den formulierten Fragen der EDK orientiert. Zum anderen haben wir, aufgrund von Überschneidungen zwischen den Blöcken A1 und A2 beziehungsweise B1 bis B3 entschieden, die Empfehlungen entlang von thematischen Bereichen zu formulieren. Darüber hinaus ist zu erwähnen, dass die einzelnen Empfehlungen eng miteinander verwoben sind. Dies bedeutet, dass die Weiterverfolgung einer Empfehlung (z.B. in Bezug auf die Festlegung der Funktionen der ÜGK) auch Konsequenzen für die Weiterentwicklung der ÜGK in einem anderen Themenbereich hat (z.B. in Bezug auf die Festlegung der Stichprobenstrategie). Insbesondere sind jene Empfehlungen, die in 6.1 „ÜGK im Gesamtsystem“ formuliert sind, als zentrale Rahmung für die nachfolgenden Empfehlungen zu verstehen. Die Empfehlungen orientieren sich an der folgenden thematischen Struktur:

- Kapitel 6.1: ÜGK im Gesamtsystem
- Kapitel 6.2: Grundlagen der ÜGK
- Kapitel 6.3: Zielvorgaben der ÜGK
- Kapitel 6.4: Design der ÜGK
- Kapitel 6.5: Institutionelle Gestaltung der ÜGK

6.1 ÜGK im Gesamtsystem

6.1.1 Weiterführung der ÜGK bei substanziellen Anpassungen im Gesamtdesign, der institutionellen Verankerung und Gestaltung sowie in der Zuweisung von finanziellen und personellen Ressourcen

Die Gesamtsicht der vorhandenen Daten zeigt deutlich, dass das Ziel der „Überprüfung des Erreichens der Grundkompetenzen“ im Sinne einer Systemevaluation auf nationaler Ebene breit akzeptiert ist. Ebenso wird die aktuelle Auswahl der getesteten Fachbereiche sowie die Auswahl der Jahrgangsstufen mit einer sehr hohen Zustimmung als sinnvoll erachtet, mit einem noch deutlicheren positiven Votum für Schulsprache und Mathematik (im Vergleich zu Fremdsprachen und Naturwissenschaften) und für die Überprüfung des Erreichens der Grundkompetenzen am Ende des 8. und 11. Schuljahres HarmoS (im Vergleich zum 4. Schuljahr).
Die Analysen zeigen aber ebenso deutlich, dass es grosse Unterschiede in den Erwartungen und Bedürfnissen der Kantone in Abhängigkeit der lokal und regional vorhandenen Qualitätsentwicklungsstrategien und Datengrundlagen gibt. Darüber hinaus kann zwar in Bezug auf die institutionelle Verankerung und Umsetzung der ÜGK eine deutliche Verbesserung im Vergleich zur Startphase beobachtet werden, allerdings sind noch verschiedene Herausforderungen zu bearbeiten. Ebenso verdeutlichen die Ergebnisse klar, dass der Nutzen der ÜGK (auch im Vergleich zum Aufwand) für die Identifikation von Stärken und Schwächen des Bildungssystems in Kantonen oder Bildungsregionen sowie für die Qualitätsentwicklung noch eingeschränkt ist, so dass das Potenzial der ÜGK für die Systemevaluation und die Qualitätsentwicklung des Bildungssystems noch nicht optimal genutzt wird. Besonders zentral ist dabei, dass bislang nicht deutlich geworden ist, was denn unter „Harmonisierung“ im Zusammenhang mit der Überprüfung des Erreichens der Grundkompetenzen verstanden beziehungsweise woran diese gemessen wird.
Die Gutachtengruppe empfiehlt, die ÜGK mit einer langfristigen Perspektive weiterzuverfolgen, jedoch mit substanziellen Anpassungen im Gesamtdesign, der institutionellen Verankerung und Gestaltung sowie in der Zuweisung von finanziellen und personellen Ressourcen.

6.1.2 Langfristige Planung bezüglich Erhebungsdesign gewährleisten und Durchführung vertraglich verbindlich absichern

Ein zentraler Kritikpunkt, der durchgängig formuliert wird, und den die Gutachtengruppe teilt, ist die bislang ungenügende Planungssicherheit. Dies verhindert in kritischem Masse, dass die ÜGK als Instrument für die nationale Systemevaluation effektiv werden kann. Zwar können zum Erreichen der Grundkompetenzen auf nationaler und kantonaler Ebene punktuell gewisse Aussagen gemacht werden, aber die Qualität und Aussagefähigkeit der Erhebungen sind durch die Kurzfristigkeit der Planung gefährdet. Zum einen würde mehr Kontinuität in der Entwicklung von Aufgaben, Tests und Kontextfragebogen vermutlich bessere Instrumente ermöglichen (z.B. ein Fragebogen, der über die Jahrgangsstufen und die Fachbereiche hinaus konzeptionell belastbar und umsetzbar ist); zum anderen würden Daten entstehen, die auch etwas zur Qualitäts*entwicklung* aussagen und damit der Bildungspolitik mehr Hinweise für Stärken und Veränderungsbedarf geben können.
Von besonderer Bedeutung ist die Planungssicherheit zudem für die bildungspolitischen Diskussionen und Strategien in den Kantonen und Sprachregionen, die ihr eigenes Monitoringsystem in Relation zu den ÜGK entwickeln beziehungsweise umsetzen können müssen. Ein längerer Planungshorizont stärkt die Relevanz der ÜGK im öffentlichen Diskurs. Darüber hinaus ist Planungssicherheit absolut zentral in Bezug auf das Gewinnen und Halten von Fachexpert*innen, ohne die die ÜGK nicht unter Berücksichtigung der wissenschaftlichen Standards realisiert werden kann. Für eine langfristige, nachhaltige Implementation der ÜGK braucht es somit einen klaren Auftrag, verbunden mit angemessenen Ressourcen und vertraglicher Sicherheit für jene Institutionen und Personen, die die ÜGK umsetzen.

Die Gutachtengruppe empfiehlt, eine langfristige Planung bezüglich Erhebungsdesign zu gewährleisten und die Durchführung vertraglich verbindlich abzusichern.

6.1.3 Verhältnis der ÜGK zu Monitoringinstrumenten in den Kantonen und Sprachregionen klären

In den Kantonen und Sprachregionen sind in unterschiedlicher Dichte eigene Monitoring- und Qualitätsentwicklungsinstrumente implementiert. Der wahrgenommene Nutzen der ÜGK in den Kantonen hängt von dieser lokalen Datenqualität und -infrastruktur sowie von der fachlichen Expertise in den Kantonen und Sprachregionen ab. Je nachdem erhält die ÜGK eine *ergänzende Funktion* (wenn im eigenen Kanton eine umfangreiche Datendichte und -qualität vorhanden ist) oder eine *Ersatzfunktion*, wenn das kantonale Bildungsmonitoring zentral von der Durchführung der ÜGK abhängt.
Um den unterschiedlichen Bedürfnissen der Kantone gerecht zu werden, **empfiehlt die Gutachtengruppe, eine modulare Struktur der ÜGK mit einem Kernmodul und einem Zusatzmodul festzulegen, wobei die Ergebnisse beider Module im nationalen Bericht dokumentiert werden sollten.**
Die genaue Ausgestaltung des Zusatzmoduls, welches aus Kapazitätsgründen für alle interessierten Kantone gleich sein müsste, könnte am besten im Gespräch zwischen dem ICER und den interessierten Kantonen entwickelt werden. Keinesfalls wird empfohlen, die ÜGK dergestalt zu modularisieren, dass einzelne Lehrpersonen und Schüler*innen eine eigene fachliche Rückmeldung zu ihren Leistungen erhalten, so dass sie darauf aufbauend im Sinne eines direkten Feedbacks ihren Unterricht oder das Lernen reflektieren können. Ein Systemevaluationsinstrument muss strikt von Instrumenten der individuellen Kompetenzdiagnostik getrennt werden, da für die wissenschaftlichen Voraussetzungen fehlen, diese reliabel und valide realisieren zu können.
Die Gutachtengruppe empfiehlt, die Frage zu klären, mit welchen kantonalen Instrumenten und auf welcher Ebene (z.B. auf Ebene Schule) eine Verknüpfung mit den Daten der ÜGK sinnvoll und gewünscht ist.
Wichtig dabei ist, dass die Klärung dieser Fragen nicht nur technisch (psychometrisch), sondern auch aus fachlich-inhaltlicher Sicht sehr anspruchsvoll ist. So beruhen beispielsweise die kantonalen Leistungstests und die Instrumente der ÜGK kaum auf gleichen Konzepten. Dadurch ist die Gefahr der Überlappung zwar sehr gering, aber ebensowenig können Synergien zwischen den verschiedenen Instrumenten genutzt werden. Beispielsweise würde der Nachweis positiver Zusammenhänge zwischen unterschiedlichen Instrumenten beide ein Stück weit validieren. Dieses grosse Feld zu bearbeiten, ist somit nur mit grossem zusätzlichem Aufwand leistbar – falls überhaupt.

6.1.4 Aufgaben und Rollen von Politik, Bildungsadministration und Forschung/Wissenschaft im Gesamtgefüge der ÜGK klären

Entsprechend den Analysen zu den implementierten Prozessen und Strukturen der ÜGK gibt es verschiedene Hinweise, dass die Aufgaben und Rollen von Politik, Bildungsadministration und Forschung/Wissenschaft noch nicht optimal festgelegt sind. Dies zeigt sich beispielsweise in der formalen Verortung der ADB, bei der Festlegung der Schwellenwerte, bei der Entwicklung des Kontextfragebogens, beim Datennutzungskonzept oder bei der Vorstellung, dass die Forschenden ein intrinsisches Interesse haben sollten, ihr Fachwissen für die Realisierung der ÜGK einzubringen, was zur Folge hat, dass (in Teilen) Forschende weniger als Dienstleister*innen, sondern als Kooperationspartner*innen betrachtet werden, die somit in gewissem Ausmass Eigenleistungen erbringen sollten. Diese Beispiele verweisen auf ein Spannungsfeld zwischen der ÜGK als politisches Systemevaluationsinstrument und der ÜGK als (mögliches) Forschungsinstrument.
Die Gutachtengruppe empfiehlt, dieses Spannungsfeld zu klären und bei der Gesamtstruktur, aber auch bei den einzelnen Teilprozessen – zum Beispiel bei der Entwicklung des Kontextfragebogens und der formalen Verortung der ADB – klare Aufgabenteilungen einzuführen und diese rechtlich zu verankern.

Dabei sollte berücksichtigt werden, dass die ÜGK ein bildungspolitisches Systemevaluationsinstrument ist. Die bildungspolitischen Akteur*innen legen die Ziele, das Erhebungsdesign und die Erhebungsinstrumente im Sinne eines strategischen Entscheids fest (z.B. welche Themen im Kontextfragebogen integriert werden sollen). Die Eigeninteressen der Forschenden sind dem politischen Interesse nachgelagert. Im Gegenzug muss die Umsetzung der politischen Entscheide nach wissenschaftlichen Standards und ohne politischen Einfluss erfolgen (z.B. anhand welcher Skalen und Items diese Themen operationalisiert werden können). Dabei sind die Forschenden Dienstleister*innen, die mittels entsprechender Aufträge finanziert werden. In Bezug auf die Entwicklung des Kontextfragebogens bedeutet dies, dass – analog zur Beschreibung der Grundkompetenzen und deren Ausformulierung für die jeweilige Testentwicklung – ein Rahmenkonzept der Befragung der Schüler*innen entwickelt werden sollte, ähnlich dem Questionnaire Framework bei PISA. Dieses Rahmenkonzept könnte in einem mehrstufigen Dialog zwischen Wissenschaft (wissenschaftliches Konsortium, vertreten durch ICER) und Politik (Kosta HarmoS und EDK) entstehen. Die Wissenschaft kann mögliche Fragestellungen und Themenfelder vorschlagen, die von der Politik priorisiert werden; die operative Ausgestaltung nach dem State of the Art der Sozialforschung wäre dann eindeutig Aufgabe der Wissenschaft.
Wenn anhand des politisch festgelegten Erhebungsdesigns auch wissenschaftlich interessante Fragestellungen bearbeitet werden können, wäre dies als Zusatznutzen zu begrüssen und dort, wo bildungspolitische mit wissenschaftlichen Interessen kompatibel sind, idealerweise umzusetzen. Durch die Ausarbeitung eines Datennutzungskonzeptes ist eine vielversprechende Grundlage geschaffen, dass die ÜGK auch als Forschungsinstrument genutzt werden kann. So sollen – etwa im Unterschied zu Deutschland – Längsschnittstudien und Verknüpfungen mit anderen Datensätzen auf Individualebene ausdrücklich ermöglicht werden. Dies stellt aus Sicht der Gutachtengruppe einen potenziell starken Anreiz für relevante Forschung dar. Allerdings bleiben im Beschlussdokument die genauen Kriterien und Verfahren für eine Genehmigung des Datenzugangs etwas im Dunkeln.
Die Gutachtengruppe empfiehlt, die wissenschaftliche Nachnutzung der aufbereiteten Datensätze innerhalb der Scientific Community noch stärker zu bewerben und das Antrags- und Genehmigungsgeschehen systematisch zu evaluieren und zu vereinfachen. Es ist zu überlegen, ob für Konfliktfälle zwischen den Interessen der Politik und den Interessen der Forschung ein geregeltes Clearing-Verfahren konzeptioniert werden sollte.

6.2 Grundlagen der ÜGK

6.2.1 Funktionen der ÜGK klären

Gemäss aktuellen Vorgaben hat die ÜGK die Funktion der Systemevaluation auf nationaler Ebene. In den Gesprächen und in der standardisierten Befragung wird aber mehrfach darauf hingewiesen, dass die ÜGK noch weitere Funktionen zu übernehmen habe, insbesondere auch das Bereitstellen von Erklärungswissen („Warum gibt es die identifizierten Unterschiede?“) oder Handlungswissen für Politik und Schulpraxis („Was können wir tun?“).
Ein gut belegter Befund ist, dass ein Systemevaluationsinstrument nicht gleichzeitig ein Instrument für die Schulpraxis sein kann (z.B. als Diagnoseinstrument in der Hand einzelner Lehrpersonen), weil dies zu Widersprüchen in der Handhabung des Instruments führt und keinen der Zwecke gut erfüllen lässt. Auch das Potenzial, unmittelbare Hinweise für die Veränderung der Schulpraxis zu generieren, ist in der Regel begrenzt. Allerdings ist es möglich, durch eine kluge Kombination von Stichprobendesign und Erhebungsdesign, auch Erkenntnisse für die Erklärung von Leistungsunterschieden zwischen Kantonen und Sprachregionen zu erhalten. Dazu ist es notwendig, weitergehende „Kontext“informationen zu erfassen, so auch zur Qualität der Lernumgebungen der Schüler*innen, und Einschätzungen der Lehrpersonen und Schulleitungen oder Informationen durch die Eltern zu berücksichtigen.
Die Gutachtengruppe empfiehlt, die Funktionen der ÜGK zu klären und allenfalls das Erhebungsdesign so zu erweitern, dass mittels der Daten nicht nur Unterschiede zwischen den Kantonen und Sprachregionen festgestellt werden können, sondern dass auch die Möglichkeit besteht, diese Unterschiede empirisch detaillierter zu untersuchen und Erklärungsansätze zu entwickeln.

6.2.2 Frage nach den Grundkompetenzen klären

Während bei den Befragten ein deutlicher Konsens besteht, die ÜGK auf Grundkompetenzen auszurichten, ist sichtbar geworden, dass meist unklar ist, wie viele der Schüler*innen denn „grundkompetent“ sein müssen, damit die Kantone mit dem Ergebnis zufrieden sind oder nicht. Grundkompetenzen, die auf Minimalstandards beruhen, wie sie im Rahmen der ÜGK festgelegt sind, bedeuten im Prinzip, dass alle Schüler*innen diese Grundkompetenzen erreichen müssen. Für das Etablieren eines nachhaltigen Bildungsmonitorings, das tatsächlich auch dazu dient, Schwächen im Bildungssystem zu identifizieren, ist es somit zentral festzulegen, ab wann Handlungsbedarf (auf nationaler Ebene oder in den einzelnen Kantonen) besteht, zumal es – vermutlich – kaum möglich sein wird, das Ziel des Erreichens der Grundkompetenzen für alle Schüler*innen zu realisieren.
Bislang besteht der Eindruck, dass in der Diskussion der Ergebnisse das (Nicht-)Erreichen des Mindeststandards nicht im Zentrum war, sondern stattdessen eher durch den Vergleich mit anderen Kantonen relativiert wurde.
So scheinen vor allem Kantone, die unter dem kantonalen Durchschnitt lagen, aufgrund der Ergebnisse Massnahmen für das Bildungssystem entwickelt zu haben; lagen hingegen die Kantone im schweizerischen Durchschnitt oder darüber, wurden häufig die Resultate zur Kenntnis genommen und dann ad acta gelegt. Eine Orientierung an einem Mindeststandard würde aber bedeuten, dass das Ausmass des Erreichens der Grundkompetenzen nach spezifischen Kriterien festgelegt wird, und dies zunächst unabhängig vom sozialen Vergleich mit anderen Kantonen (z.B. mindestens x % der Schüler*innen müssen die Grundkompetenzen erreichen). Falls dieses Kriterium nicht erreicht wird, wären entsprechende Massnahmen notwendig, auch wenn der Kanton im Durchschnitt oder sogar darüber liegen würde.
Angesichts der herausragenden Relevanz der Frage nach dem Erreichen der Grundkompetenzen und ihrer Klärung für die Weiterentwicklung der ÜGK sowie angesichts des – sogar reglementarisch – festgelegten Wunsches, nicht den sozialen Vergleich und ein mögliches Ranking zwischen den Kantonen ins Zentrum zu rücken (EDK, 2014), empfiehlt die Gutachtengruppe somit, diese Frage nach den Grundkompetenzen prioritär zu klären.

6.2.3 Frage nach der Harmonisierung der Grundkompetenzen klären

In Bezug auf das Thema der Harmonisierung verweisen die Ergebnisse der Analysen darauf, dass unklar ist, worauf sich der Anspruch auf Harmonisierung, wie er im Bildungsartikel formuliert ist, bezieht und welche weitergehenden Harmonisierungsziele die EDK erreichen möchte. Idealtypisch sind deren drei denkbar: Harmonisierung des Anspruchs, Harmonisierung des Monitorings und Harmonisierung der Ergebnisse.

Gemäss Bildungsartikel ist die *Harmonisierung des Anspruchs* zentral. Dieses Harmonisierungsziel kann mit der Formulierung gemeinsamer Bildungsstandards und Grundkompetenzen als erfüllt betrachtet werden. Eine über den Bildungsartikel hinausgehende *Harmonisierung des Monitorings* ist einerseits durch den Entscheid der EDK, das Erreichen der Grundkompetenzen zu überprüfen und andererseits mit der Durchführung der ÜGK-Erhebungen ebenfalls gegeben.

Unklar ist allerdings, inwiefern eine *Harmonisierung der Ergebnisse* angestrebt wird, anhand welcher Kriterien die Ergebnisse im Sinne der Zielerreichung positiv interpretiert werden können oder anhand welcher Indikatoren sich der Weg zu einer grösseren Harmonisierung abbilden lassen. So ist beispielsweise nicht nur zu klären, wie viele Schüler*innen denn grundkompetent sein müssen, sondern auch, wie gross die Unterschiede zwischen den Kantonen sein dürfen, damit eine Harmonisierung der Ergebnisse auf nationaler oder sprachregionaler Ebene vorhanden ist, und wie genau dieser Unterschied beziffert wird. Prinzipiell lassen sich diese Fragen auch innerhalb eines Kantons stellen, beim Vergleich etwa zwischen Gemeinden oder zwischen Stadt und Land.

Die Gutachtengruppe empfiehlt zu klären, anhand welcher Kriterien die Ergebnisse der ÜGK positiv zu interpretieren sind oder anhand welcher Indikatoren sich der Weg zu einer grösseren Harmonisierung abbilden lässt.

6.3 Zielvorgaben der ÜGK

6.3.1 An einer Orientierung an Mindeststandards und der Überprüfung des Erreichens der Grundkompetenzen festhalten sowie die 2011 formulierten Grundkompetenzen weiterhin als Basis der Aufgabenentwicklung akzeptieren

Die Gutachtengruppe empfiehlt, weiterhin an der Bestimmung des Anteils des Erreichens/Nicht-Erreichens der Grundkompetenzen festzuhalten, die Berichte zur ÜGK (wie schon 2016 und 2017) ausschliesslich darauf auszurichten und auf weiter ausdifferenzierte Kompetenzstufenmodelle zu verzichten. Vier Gründe sprechen dafür:

1. Die Grundkompetenzen sowie der Fokus auf Mindeststandards werden allgemein akzeptiert.
2. Auch der jetzige Modus der Festlegung auf Grundkompetenzen und die empirische Überprüfung des Erreichens ermöglicht es anhand der breit variierenden Testscores, das Spektrum der gemessenen Leistungen differenzierter als ausschliesslich mit dem Indikator „Grundkompetenzen erreicht/nicht erreicht“ zu analysieren.
3. Ein Wechsel auf weiter ausdifferenzierte Kompetenzstufen über den gesamten Leistungsbereich hätte zur Folge, dass der gesamte ÜGK-Prozess (Definierung von Kompetenzen auf weiteren Stufen, Definierung von Schwellenwerten für die weiteren Stufen sowie Entwicklung entsprechender Testaufgaben) neu konzipiert werden müsste, was zu erheblichen Verzögerungen führen würde.
4. Akteur*innen aus der ADB haben Zweifel geäussert, ob eine Angleichung aller Erwartungsniveaus über die Sprachregionen hinweg, mit ihren zum Teil grossen fachdidaktischen Traditionsunterschieden, überhaupt möglich wäre; der Fokus auf Mindeststandards erleichtert daher die Harmonisierung sehr.

Dessen ungeachtet wird empfohlen, die Grundkompetenzen Mathematik zu überprüfen und allenfalls den Erwartungshorizont zu adjustieren. So wurden diese möglicherweise zu anspruchsvoll formuliert. Das Audit von Fischbach und Ugen (2018) schlug daher eine Revision vor, welche die EDK an Kosta HarmoS delegiert hat. Sie wurde nach unserer Kenntnis noch nicht in Angriff genommen.

Weitere Empfehlungen

- Die Aufgabenentwicklung soll ein durchaus breit gefächertes Spektrum von Aufgabenschwierigkeiten anpeilen, auch deutlich unterhalb und oberhalb der Grenze zur Grundkompetenz, wobei um den Schwellenwert herum mehr Aufgaben zu platzieren sind als in anderen Abschnitten des Spektrums. Zudem sollte bei der Aufgabenentwicklung und -pilotierung explizit eingeplant werden, dass Aufgaben sowohl unterhalb als auch oberhalb des Schwellenwertes zur Veröffentlichung in der Praxis benötigt werden, um Grundkompetenzen zu illustrieren und von anderen Leistungsniveaus abzugrenzen.
- Jeder ÜGK-Test sollte mit einem Framework begründet werden, welches das Kompetenzkonstrukt auf fachdidaktischer Basis aktualisiert, Beispielaufgaben benennt und so den fachdidaktischen Diskurs – auch mit Lehrpersonen in der Praxis – vorantreibt. Dieses Framework sollte Ergebnisse der umfangreichen Pilotierung der Aufgaben nutzen und parallel zum jeweiligen Ergebnisbericht publiziert werden. Dies macht insbesondere den gesetzten Schwellenwert transparent: Anhand von Beispielaufgaben und Aufgabenmerkmalen soll dargestellt werden, was „grundkompetente“ Schüler*innen können, und welche Anforderungen über die Grundkompetenzen hinausgehen.
- Die EDK sollte zudem Validierungsstudien in Auftrag geben, gegebenenfalls in Zusammenhang mit Pilotierungen. Diese Validierungsstudien sollten sowohl die Scores auf der kontinuierlichen Testskala verwenden als auch die dichotomisierten Ergebnisse (0 = erreicht/1 = nicht erreicht), sodass Forschende und politische Entscheidungsträger*innen mehr Sicherheit bekommen in der Interpretation der ÜGK-Befunde.
- Psychometriker*innen, Fachdidaktiker*innen und Bildungsforscher*innen aus ÜGK-Projekten sollten in einem Grundlagenpapier fachübergreifend darstellen, was „Messung von Grundkompetenzen“ genau bedeutet.

- Die Aufgabenentwickler*innen sollten den zusätzlichen Auftrag erhalten, gemeinsam mit den Psychometriker*innen der ADB systematische Aufgabenanalysen vorzunehmen, das heisst Anforderungsmerkmale zu spezifizieren und deren Auswirkung auf das Schwierigkeitsniveau und auf differenzielle Schwierigkeitswerte in Sprachregionen oder Kantonen zu ermitteln. Dazu sollen primär die Daten der Pilotierungen verwendet werden. Ergebnisse sollen bei der Schwellenwertsetzung genutzt, in die Ausbildung von Lehrpersonen und in den fachdidaktischen Diskurs hineingetragen werden. Diese Arbeiten stellen eine Dienstleistung im Kontext der Testentwicklung dar, sind damit Teil von Serviceverträgen und erfordern entsprechende Ressourcen.
- Die vollständigen Testdaten (Antworten auf Item-Niveau und Testscores als „Plausible Values“) sollten für Sekundäranalysen zur Verfügung gestellt werden. Die Testscores in der gesamten Varianzbreite können allenfalls im schweizerischen Bildungsbericht genutzt werden, um Zusammenhänge mit nicht-kognitiven Merkmalen sowie familialen und schulischen Faktoren zu untersuchen. Der Erstbericht zur ÜGK sollte sich jedoch – sofern die politische Zielsetzung nach HarmoS unverändert bleibt – weiterhin darauf beschränken, Anteile derjenigen zu berichten, welche den Schwellenwert überschreiten, und diese Anteile mit anderen Variablen zu verknüpfen.

6.3.2 Frage klären, welche Bedeutung der soziale Vergleich zwischen den Kantonen hat und adäquate Lösungen für den Kantonsvergleich finden

Sowohl in den bisherigen nationalen Berichten wie auch in den Gesprächen wird deutlich, dass ein „sozialer“ Vergleich der Kantone in der Erreichung der Grundkompetenzen für die Befragten eine zentrale Bedeutung hat. In den nationalen Berichten wird der Vergleich mit unterschiedlichen Auswertungen realisiert.

Die Gutachtengruppe empfiehlt, die weitgehend deskriptive Ausrichtung der Berichte auf Anteile „grundkompetenter“ Schüler*innen nach Kantonen und Teilgruppen fortzusetzen. Sogenannte adjustierte Vergleiche sollten hingegen vertiefenden Publikationen vorbehalten bleiben.

So kann in Bezug auf das Erreichen der Grundkompetenzen in den Kantonen diese Adjustierung auch zu Problemen führen, da sie dazu verleitet, ein mangelndes Erreichen des gewünschten Ergebnisses im eigenen Kanton beispielsweise mit den geringeren familiären Bildungsressourcen oder der Häufung von Schüler*innen mit Migrationshintergrund zu erklären und Differenzen zwischen den Schüler*innen als „naturgegeben“ anzuerkennen. Gerade das Setzen von Mindeststandards beinhaltet aber die Idee, dass (fast) allen Schüler*innen durch das Bildungssystem ermöglicht werden soll, diese Ziele zu erreichen, unabhängig von familiären Merkmalen.

6.4 Design der ÜGK

6.4.1 An der Wahl der Jahrgangsstufen und der Fächer / Fachbereiche sowie an einer standardisierten Erhebung über alle Kantone hinweg festhalten

Wie bereits erwähnt, besteht ein breiter Konsens bezüglich der Auswahl der Fächer/Fachgruppen sowie der Jahrgangsstufen, am Ende derer die Grundkompetenzen überprüft werden sollen.
Die Gutachtengruppe empfiehlt, die Erhebungen auf jeden Fall für Mathematik und Sprachen (Schulsprache, in den höheren Jahrgangsstufen gekoppelt mit der ersten Fremdsprache) fortzuführen.
Ob Naturwissenschaften einbezogen werden können, gegebenenfalls in einer Erhebung gemeinsam mit der Mathematik, sollte nach sorgfältiger Prüfung der verfügbaren Ressourcen (von der Expertise in der Aufgabenentwicklung bis zu Ausstattung und Personal für Experimentiertests, wie sie in der HarmoS-Validierungsstudie verwendet wurden) im Rahmen einer systematischen Gesamtplanung entschieden werden. Im Anschluss an die ÜGK 2024, die Mathematik und Schulsprache in der 4. Klasse testet, sollte auch noch einmal geprüft werden, wie belastbar Untersuchungen in diesem Alter sind. Gerade das Wegfallen der 4. Jahrgangsstufe hätte aber zur Folge, dass Kohortenlängsschnittdesigns nicht mehr gut realisiert werden könnten. Für ein nationales Systemevaluationsinstrument zwingend notwendig hingegen ist eine zeitliche und inhaltliche Standardisierung der Erhebungen über alle Kantone hinweg.
Die Gutachtengruppe empfiehlt, die Erhebungen in allen Kantonen zur gleichen Zeit und mit der gleichen inhaltlichen Ausrichtung durchzuführen.

6.4.2 Erhebungsdesign klären: Trend-, Kohorten- und Längsschnitterhebungen?

Bislang wurden zwei ÜGK-Erhebungen durchgeführt, eine weitere Erhebung musste aufgrund der COVID-19-Pandemie verschoben werden. In Bezug auf die weitere Sequenzierung und inhaltliche Ausrichtung der einzelnen Befragungen stellt sich somit die Frage, in welchem zeitlichen Abstand, in welcher Jahrgangsstufe und in welchem/welchen Fachbereich(en) die weiteren Erhebungen durchgeführt werden sollen. Hierzu sind folgende grundsätzlich möglichen Erhebungsdesigns zu klären, da sie die Beantwortung von je unterschiedlichen Fragestellungen ermöglichen: Trenddesign, Kohortenlängsschnittdesign und Individual-Längsschnittdesign.[21]
Trenddesign: Trenderhebungen sind unter Bildungsmonitoringperspektive sehr wichtig. Dabei werden im regelmässigen Abstand (alle 3 bis 6 Jahre) die Leistungen der Schüler*innen einer bestimmten Jahrgangsstufe erfasst und mit den Leistungen der Schüler*innen gleicher Jahrgangsstufe in den nachfolgenden Erhebungen über die Jahre verglichen. Dies entspricht beispielsweise dem Design der PISA-Studien.
Kohortenlängsschnittdesign: Dieses Erhebungsdesign wird bei TIMSS (Trends in International Mathematics and Science Study) oder beim Bildungspanel (NEPS) verwendet. Dabei wird eine Kohorte, beispielsweise die Kohorte der Schüler*innen in der vierten Klasse zum Zeitpunkt X, erneut im achten Schuljahr zum Zeitpunkt X+4 und gegebenenfalls im elften Schuljahr zum Zeitpunkt X+7 befragt. Im Gegensatz zum „Individual-Längsschnittdesign“ (siehe nächster Punkt) wird dabei jedesmal eine neue Stichprobe gezogen. Denkbar sind ebenfalls Mehr-Kohorten-Designs, indem beispielsweise zusätzlich zur Kohorte mit Start im Jahr X eine weitere mit Start im Jahr X+6 untersucht wird. Um Erkenntnisse für das Bildungsmonitoring zu gewinnen, ist ein langfristiger Zeitplan unerlässlich.
Individual-Längsschnittdesign: Individuelle Längsschnitterhebungen sind sehr komplex, ermöglichen es aber am besten, Erklärungen für Bildungsverläufe und mögliche Differenzen in der Entwicklung der Kompetenzen der Schüler*innen im kantonalen Vergleich zu gewinnen. Sie beinhalten zusätzlich zum

[21] Je nach Erhebungsdesign muss der Umgang mit den bestehenden Testaufgaben geklärt werden. In einem Trenddesign dürfen beispielsweise die zentralen Testaufgaben nicht veröffentlicht werden, damit sie in den nachfolgenden Trenderhebungen weiterverwendet werden können.

Kohortenlängsschnittdesign die Verfolgung der einzelnen Schüler*innen im Bildungsverlauf. Individuelle Längsschnitterhebungen sind darüber hinaus für Fragen der Bildungsforschung sehr zentral, bedingen aber eine Beobachtung der einzelnen Schüler*innen anhand eines Identifikationscodes über die Bildungslaufbahn hinweg. Angesichts der verschiedenen Klassen-, Schul- und Kantonswechsel ist dies sehr voraussetzungsreich.

Die Gutachtengruppe empfiehlt, das ÜGK-Erhebungsdesign im Kern als Trenddesign zu konzipieren, da dies die Grundvoraussetzung für ein systematisches langfristiges Bildungsmonitoring ist. Inwiefern weitergehende Designs gewünscht und unter welchen Voraussetzungen sie möglich sind, müsste im Austausch aller relevanten Akteur*innengruppen geklärt werden.

6.4.3 Inhalte der nationalen Berichte klären

In der Berichterstattung der Ergebnisse erfolgte bislang eine Trennung zwischen den zentralen Ergebnissen zum Erreichen der Grundkompetenzen, die im nationalen Bericht publiziert werden, und den Vertiefungsstudien, die die Kantone in Auftrag geben können, um kantonal interessante Fragestellungen zu untersuchen. In den Äusserungen der von uns Befragten wird allerdings moniert, dass die Auswertungen, die im nationalen Bericht publiziert sind, die Informationen aus dem Kontextfragebogen zu wenig berücksichtigten, der grosse Aufwand für die Schulen und kantonale Systeme somit nur zu umfangmässig geringen Erkenntnissen führe. Zum anderen haben einzelne Kantone Mühe, Vertiefungsstudien zu finanzieren und die Dauer, bis diese Analysen vorliegen, wird als zu lang kritisiert.

Die Gutachtengruppe empfiehlt, das Konzept der Kernauswertungen und der Vertiefungsstudien zu überdenken sowie die Kontextinformationen, die bedeutsam sind für die Interpretation der Befunde, systematisch in die Kernanalysen einzubeziehen und damit den nationalen Bericht auszubauen.

6.4.4 Kontextinformationen erweitern, in Abhängigkeit der Funktionen der ÜGK

Im Zusammenhang mit der Frage nach der Identifikation möglicher Einflussfaktoren auf die Leistungen der Schüler*innen können im jetzigen System einzig individuelle Faktoren der Schüler*innen (z.B. Geschlecht, familiärer Hintergrund) berücksichtigt werden, da parallel zu den Leistungstests lediglich ein Fragebogen für die Schüler*innen eingesetzt wird. Bisherige nationale und internationale Forschungsergebnisse zeigen aber deutlich, dass Unterschiede in den Leistungen der Schüler*innen nicht nur durch diese Merkmale der Schüler*innen erklärt werden können, sondern dass weitere Faktoren, insbesondere der Unterricht und die Kompetenzen der Lehrpersonen, aber auch schulische Variablen (z.B. Kooperation im Kollegium, Schulleitungshandeln) bedeutsam sind. Diese sind zudem die wichtigsten Anknüpfungspunkte für die Qualitätsentwicklung. Darüber hinaus tragen insbesondere bei den jüngeren Schüler*innen Informationen durch die Eltern wesentlich zum besseren Verständnis der Unterschiede zwischen den Leistungen der Schüler*innen bei.

Die Gutachtengruppe empfiehlt, den konzeptuellen Rahmen des Kontextfragebogens in den HarmoS-Gremien grundsätzlich zu diskutieren, wenn Einigkeit über die Zielsetzung und die Grundlagen des Stichprobendesigns besteht.

- Bleibt es bei einer blossen Deskription des Anteils von Schüler*innen pro Kanton, die die Grundkompetenzen erreichen, werden sinnvollerweise nur diejenigen Informationen erfragt, die für das sogenannte „Hintergrundmodell“ der Schätzung von Testscores (Plausible Values) benötigt werden.
- Zielt die ÜGK jedoch auf weitergehende Fragen, empfiehlt die Gutachtengruppe, den Einsatz zusätzlicher Erhebungsinstrumente für Lehrpersonen, Schulleitungen und Eltern zu prüfen, damit ein besseres Verständnis für mögliche Unterschiede zwischen den Kantonen gewonnen werden kann. Dies wäre auch im Einklang mit dem State of the Art von international vergleichenden Studien.

6.4.5 Stichprobenziehung entsprechend den Funktionen der ÜGK anpassen

Entsprechend der aktuellen Zielsetzung der ÜGK, die Grundkompetenzen der Schüler*innen im Sinne eines Systemmonitorings auf nationaler Ebene zu überprüfen, ist die Stichprobenziehung in Bezug auf eine repräsentative Auswahl der Schüler*innen pro Kanton funktional. Sobald allerdings Erklärungswissen generiert werden soll, welches mehr als den Einfluss individueller Merkmale wie Geschlecht oder den familiären Hintergrund der Schüler*innen einzubeziehen hat, ist eine Anpassung des Stichprobendesigns notwendig. Denkbar sind hierzu zwei Ansätze: Stichprobenziehung auf Klassenebene oder Stichprobenziehung auf Klassen- und Schulebene, falls nicht nur Klassenvariablen, sondern auch Schulvariablen als mögliche Erklärungsindikatoren untersucht werden sollen.
Die Gutachtengruppe empfiehlt, die Stichprobenziehung entsprechend den Funktionen der ÜGK anzupassen.

6.4.6 Öffentlichen Kommunikation der Ergebnisse der ÜGK verbessern

Die Gutachtengruppe empfiehlt, die (politische) Kommunikation der Ergebnisse auf nationaler und kantonaler Ebene zu verbessern und dabei insbesondere auch den Bedürfnissen der Kantone für die innerkantonale Kommunikation angemessen Rechnung zu tragen. Damit könnten die Ergebnisse der ÜGK auch in der Öffentlichkeit noch sichtbarer werden. Aus fachlicher Perspektive könnten die wissenschaftlichen Akteur*innen bei der Interpretation der Ergebnisse stärker Unterstützung anbieten. Dies wäre allerdings als verbindlicher Auftrag zu formulieren und entsprechend zu budgetieren.

6.4.7 Weitere Empfehlungen

Es wird davon abgeraten, die Piloterhebungen in die Hand der Lehrpersonen zu geben, dies insbesondere deshalb, weil Pilotierungs- und Haupttesterhebungen unter den gleichen Bedingungen durchgeführt werden müssen (z.B. beide durch externe Testleitungsadministration oder beide durch schulseitige Testadministration).
Die Technischen Berichte sollten einheitlich ausfallen, wofür ein Template sinnvoll erscheint, damit diese Berichte und Dokumentationen über alle Testdomänen und Jahrgangsstufen hinweg so weit wie möglich harmonisiert sind. Pro Erhebungszeitpunkt sollte zudem ein einziger umfassender Technischer Bericht publiziert werden.

6.5 Institutionelle Gestaltung der ÜGK

6.5.1 Klare sowie vereinfachte Zuständigkeits-, Entscheidungs- und Verantwortungsstrukturen implementieren

Aufgrund der Komplexität der ÜGK und der Notwendigkeit, Teilprozesse aufeinander abzustimmen, braucht es klare Strukturen für Entscheidungen. Derzeit ist in der konkreten Praxis eine Diffusion in Zuständigkeiten und Verantwortlichkeiten zu beobachten, die dadurch verstärkt wird, dass auch vertragliche Unklarheiten bestehen.

Dies zeigt sich zum einen in der Unklarheit, ob die Universität Bern tatsächlich „Kooperationspartnerin“ der EDK ist oder eine Institution, die einen bestimmten Auftrag der EDK als dienstleistende Institution bearbeitet. Zum andern, ob die Universität Bern beziehungsweise das ICER nur eine Koordinations- oder auch eine Leitungsaufgabe zu übernehmen hat. So soll zwar die Universität Bern gemäss der Kooperationsvereinbarung zwischen der EDK und der Universität Bern die „Gesamtleitung der Prozesse“ übernehmen und mit verschiedenen Institutionen aus allen Sprachregionen „kooperieren“ (Absatz 3.2.2)[22]. Gemäss Absatz 3.2.3 wird dann allerdings einzig der Auftrag zur Koordination von verschiedenen Bereichen festgelegt. Die Mischung zwischen Leitungs- und Koordinationsfunktionen akzentuiert sich des Weiteren dadurch, dass die Universität Bern gemäss Vereinbarung mit den verschiedenen Institutionen „Kooperationsvereinbarungen“ abzuschliessen und darin die zu übernehmenden Rollen zu definieren hat (Absatz 3.2.2). In diesem Vorgehen zeigt sich ebenfalls eine Vermengung zwischen Leitungs- und Koordinationsfunktionen.

Unklarheiten manifestieren sich auch an anderen Stellen. So stellt sich beispielsweise die Frage nach den Funktionen des „wissenschaftlichen Konsortiums“, das in verschiedenen Reglementen beschrieben wird. Zum einen scheint, dass mit dem ICER das wissenschaftliche Konsortium umgesetzt wurde, zum anderen ist gemäss Unterlagen noch ein „wissenschaftliches Konsortium Bildungsforschung und -evaluation“ in Diskussion. Momentan scheint es sich eher um ein Verbund von interessierten Institutionen und Personen zu handeln, die sich vor allem sporadisch zum Zweck des Informationsaustauschs treffen, so dass es kaum eine tragende und strategische Rolle in langfristiger Perspektive ausüben kann. Genau dies wäre aber bedeutsam, um der wissenschaftlichen Komplexität der ÜGK gerecht werden zu können.

Darüber hinaus bestehen nicht nur vertragliche Vereinbarungen zwischen der EDK und der Universität Bern, sondern auch mit der Schweizerischen Koordinationsstelle für Bildungsforschung (SKBF) zur „Domizilierung der ADB“ und der FH Graubünden zur Realisierung von für die ÜGK zentralen technischen Aufgaben. In diese Aufgabengebiete ist die Universität Bern beziehungsweise das ICER nicht involviert. Sie kann – respektive darf – dafür aktuell keine Leitungsfunktionen übernehmen, was zu einem erhöhten Koordinationsaufwand und bedeutsamen Schnittstellen-Risiken führt. Gerade die Software, die von der FH Graubünden bereitgestellt wird, hat massive Folgen für den gesamten ÜGK-Prozess, wobei die Befragten auf verschiedene Probleme in der bisherigen Praxis hingewiesen haben. Darüber hinaus bleibt unklar, warum mit der SKBF und der FH Graubünden separate Verträge bestehen und die entsprechenden Aufgaben nicht auch mittels Kooperationsvereinbarungen über die Universität Bern geregelt werden.

Die Gutachtengruppe empfiehlt, die komplexen vertraglichen Strukturen mit Parallelverträgen der EDK zu überdenken und zusammenzuführen, da ansonsten die „Einheit der Materie“ verletzt wird und wichtige Teilaufgaben und -prozesse in unterschiedlichen Verantwortungsbereichen liegen.

Zudem wird eine Koordinationsfunktion (ohne die explizite Übernahme einer Leitungsfunktion) durch die Universität Bern beziehungsweise das ICER als nicht ausreichend und langfristig dysfunktional angesehen. Es braucht Klarheit bezüglich der Entscheidungs- und Prozessstrukturen und ein eindeutiges, vertraglich festgelegtes Mandat für die Universität Bern beziehungsweise das ICER zur Gesamtleitung

[22] Kooperationsvereinbarung betreffend die Zusammenarbeit zur Durchführung der ÜGK-Erhebungen in 2020 und 2022 und zur kontinuierlichen wissenschaftlichen Koordination der Überprüfung des Erreichens der Grundkompetenzen (ÜGK)

der Prozesse in den zentralen Bereichen. Am augenfälligsten ist dies bei der Aufgabenentwicklung und der Entwicklung des Kontextfragebogens. Um ein Gesamtkonzept für einen tragfähigen Kontextfragebogen zu entwickeln, der über die Jahrgangsstufen hinweg wesentliche Kontextinformationen zum Lernen der Schüler*innen erfasst, reicht ein rein koordinierendes Vorgehen nicht. Es braucht eine klare inhaltliche Gesamtstrategie und demzufolge strategische Entscheide. Eine Gesamtleitung ist sodann bezüglich der einzelnen Teilbereiche in der Praxis so zu konkretisieren, dass die verschiedenen involvierten Institutionen und Sprachregionen ihre Aufträge bestmöglich und im Sinne der Gesamtzielerreichung realisieren können.

Die Gutachtengruppe empfiehlt, die Leitungsfunktion der Universität Bern beziehungsweise des ICER zu stärken und die Entscheidungs- und Prozessstrukturen klarer aufzubauen. Darüber hinaus sollte der Punkt „Kooperation“ zwischen der Universität Bern und der EDK geklärt und ausdifferenziert werden: Welche Aufgaben liegen in der Kompetenz der EDK und welche im Sinne eines Auftrags in der Kompetenz der Universität Bern beziehungsweise des ICER?

6.5.2 Testentwicklung als komplexes und anspruchsvolles wissenschaftliches Aufgabenfeld anerkennen, mit genügend personellen Ressourcen ausstatten und institutionell in die wissenschaftliche Projektstruktur der ÜGK einbinden

Die Aufgabendatenbank der EDK (ADB) ist eine der Kernakteur*innen im Prozess der Überprüfung des Erreichens der Grundkompetenzen. Bei den entsprechenden Testentwicklungen und psychometrischen Analysen handelt es sich um komplexe und anspruchsvolle wissenschaftliche Aufgaben, die einzig von wissenschaftlichen Fachexpert*innen unter Berücksichtigung des State of the Art in Forschungsmethodik und Fachdidaktik realisiert werden können. Ohne einen hohen Standard in der Aufgaben- und Testentwicklung können die ÜGK ihre Funktion bezüglich des Systemmonitorings nicht erfüllen und die Ziele eines systematischen Monitorings auf nationaler Ebene nicht erreichen.

Bislang ist die ADB allerdings nicht als wissenschaftliche, sondern als politische Akteurin interpretiert und formal-rechtlich positioniert worden. Damit zusammenhängend erfolgte eine formal-rechtliche Trennung der Verantwortungsträger*innen der Test- und Aufgabenentwicklung einerseits und Durchführung der ÜGK andererseits. Auch die fehlende institutionelle Anbindung der ADB an eine wissenschaftliche Institution sowie ihre isolierte Position innerhalb der ÜGK stellen ein substanzielles Risiko dar. Zudem ist die ADB personell massiv unterbesetzt und bietet qualifiziertem Personal – auch aufgrund der Kurzfristigkeit der Verträge – wenig Perspektiven. Damit besteht nicht nur ein erhebliches Risiko in Bezug auf kritischen Know-how-Verlust, sondern auch die Gefahr, dass die komplexen Anforderungen an die Testentwicklung – mehrere Jahrgangsstufen, Fächer, Sprachen – mittel- und langfristig nicht adäquat bearbeitet werden können. Dies trotz des hohen Engagements der aktuellen Fachperson(en) in der ADB.

Die EDK argumentierte in Antwort auf die im Audit-Report von Fischbach und Ugen (2018) formulierten Kritikpunkte, dass durch eine Professionalisierung der Prozesse und durch den Erfahrungsgewinn real mehr Zeit für die Entwicklung der einzelnen Aufgaben zur Verfügung stünde. Diese Entscheide können aufgrund der Komplexität der Aufgaben- und Testentwicklung in drei Sprachen nicht überzeugen.

Die Gutachtengruppe empfiehlt,

- **eine ausreichende personelle Ausstattung mit längerfristiger Perspektive, eine systematische und institutionelle Anbindung an wissenschaftliche Institutionen sowie eine Loslösung vom GS EDK und eine Einbindung in die Projektstruktur der ÜGK zu ermöglichen.**
- **die Vorschläge des Audits von Fischbach und Ugen (2018) konsequent aufzugreifen, indem die Teilprozesse von Aufgabenentwicklung, Testzusammenstellung, Testdesign, Stichprobendesign, Standard-Setting sowie Datenanalyse technisch und institutionell stärker verzahnt werden, wie dies in Large-Scale-Assessment-Programmen anderer Länder (Luxemburg, Österreich, Deutschland) der Fall ist.**

6.5.3 Fachdidaktik(en) stärker in den Gesamtprozess der ÜGK einbinden

Vertretungen der Fachdidaktiken sind systematisch ausschliesslich in der Aufgabenentwicklung und des Standard-Settings eingebunden. In dieser Rolle sind sie über die ADB an die Projektleitung ÜGK angebunden. Damit fliesst die Perspektive der Fachdidaktiken derzeit vorrangig in die Aufgabenentwicklung ein, nicht aber in andere Phasen der ÜGK, etwa in gesamtkonzeptionelle Aspekte oder die Ergebnisanalyse und -interpretation. Eine verlässliche, transparente, institutionalisierte Basis für die umfängliche Berücksichtigung der wichtigen fachdidaktischen Perspektive in der ÜGK besteht nicht.

Die Gutachtengruppe empfiehlt,

- **den begonnenen Weg einer multidisziplinären Aufgaben- und Testentwicklung zu forcieren und in allen Phasen der Entwicklung Vertreter*innen aus Fachdidaktik, Unterrichtspraxis und Edumetrie einzubeziehen. Zu diesem Zweck sollte ein integriertes und langfristiges Konzept der Aufgaben- und Testentwicklung erstellt werden, aus dem der Professionalisierungsanspruch der Aufgabenentwicklung erkennbar wird.**
- **Strukturen zu finden, die die Berücksichtigung der Perspektive der Fachdidaktiken in der Gesamtkoordination der ÜGK ermöglichen. Dies sollte systematisch erfolgen, so dass die Perspektive der Fachdidaktiken in der Gesamtkoordination gezielt einbezogen wird und dadurch eine Brücke zum Unterricht wie auch zur Aus- und Weiterbildung von Lehrpersonen geschlagen werden kann. Zudem ermöglicht dies, die derzeit als erziehungswissenschaftlich-, psychologisch- sowie soziologisch-fokussiert wahrgenommene Daten- und Ergebnisnutzung um eine fachdidaktische, unterrichts- und schulnähere Perspektive anzureichern.**

6.5.4 Sprachregionale Unterschiede in Bezug auf fachliches Lernen und Aufgaben herausarbeiten sowie für die Testentwicklung nutzen

In den Gesprächen zeigen sich verschiedene Hinweise, dass in der Entwicklung des Kontextfragebogens oder der Testaufgaben die sprachregional verschiedenen fachdidaktischen Kulturen zu wenig berücksichtigt und sichtbar geworden sind.

Die Gutachtengruppe empfiehlt, die Daten der ÜGK zu den Testaufgaben in Bezug auf allfällige Differenzen zwischen den Sprachregionen zu analysieren (bzw. soweit solche Analysen stattfinden, sichtbar zu publizieren). Dies kann zum einen als Validitätsprüfung, zum anderen aber auch als Anregung für die fachdidaktische Weiterentwicklung der (eventuell vorhandenen) unterschiedlichen Lern- und Unterrichtskulturen genutzt werden kann.

Empfohlen wird sodann, die (allfällig vorhandenen) unterschiedlichen Lern- und Unterrichtskulturen sorgfältig herauszuarbeiten, Differenzen und Überschneidungen zu identifizieren und in einem fachlichen Diskurs für die Entwicklung der Erhebungsinstrumente zu nutzen.

I. Anhang – Literaturverzeichnis

Altrichter, H., & Maag Merki, K. (2016). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, 2., erweiterte Ausgabe (S. 1-28). Springer VS.

Angelone, D., & Keller, F. (2019a). *Überprüfung des Erreichens der Grundkompetenzen (ÜGK) im Fach Mathematik im 11. Schuljahr: Technische Dokumentation zur Testentwicklung und Skalierung.* Aufgabendatenbank der EDK.

Angelone, D., & Keller, F. (2019b). *Überprüfung des Erreichens der Grundkompetenzen (ÜGK) in den Fächern Schulsprache und erste Fremdsprache im 8. Schuljahr: Technische Dokumentation zur Testentwicklung und Skalierung.* Aufgabendatenbank der EDK.

Angelone, D., Keller, F., Arnold, J., Jaun-Holderegger, B., Hauser, M., Eckstein, B., Erzinger, A., Leutwiler, L., Nell-Tuor, N., & Schönenberger, S. (2019). *Machbarkeitsstudie zur Überprüfung der Grundkompetenzen im 4. Schuljahr.* Aufgabendatenbank der EDK.

EDK (2007). *Interkantonale Vereinbarung über die Harmonisierung der obligatorischen Schule (HarmoS-Konkordat); Umsetzung auf der Ebene der interkantonalen Koordination vom 25./26. Oktober 2007.*

EDK (2014). *Organisationsreglement über die Durchführung der Überprüfung des Erreichens der Grundkompetenzen*. Bern: EDK.

EDK (2018). Kooperationsvereinbarung zwischen der EDK und der Universität Bern betreffend die Zusammenar*beit zur Durchführung der ÜGK-Erhebungen in den Jahren 2020 und 2022 und zur kontinuierlichen wissenschaftlichen Koordination der Überprüfung des Erreichens der Grundkompetenzen (ÜGK) gemäss Artikel 4 Absatz 2 des ÜGK-Reglements*. Bern: EDK.

EDK (2019). *Überprüfung des Erreichens der Grundkompetenzen: Bericht zu den Erhebungen 2016 und 2017*. Bern: Generalsekretariat der EDK.

Erzinger, A., Hauser, M., Dutrevis, M., Hascher, T., Keller, R., Lenz, P., & Soucis, A. (2019). *Erläuterungen zu den Skalen des Kontextfragebogens der ÜGK Sprachen 2017: Theoretischer Hintergrund, Inhalte und Konstrukte*. Bern & St. Gallen: Universität Bern, Pädagogische Hochschule St. Gallen, Service de la recherche en éducation (SRED), Universität Fribourg.

Fischbach, A., & Ugen, S. (2018). *ÜGK/COFO Mathematics 2016: Audit report: commissioned by EDK/CDIP upon the request of KOSTA HarmoS.* http://edudoc.ch/record/204067/files/UeGK_Audit_Report_230218.pdf

Fuchs, G. (2019). Prognosegüte bildungsstandardisierter Tests. Dissertation, FU Berlin. http://dx.doi.org/10.17169/refubium-1687

Hambleton, R.K. & Zenisky, L. (2010) Translating and adapting tests for cross-cultural assessments, in D. Matsumoto and F. Van de Vijer (eds.) Cross-cultural research methods in psychology, (pp. 46–70). Cambridge: Cambridge University Press.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise.* Bundesministerium für Bildung und Forschung.

Kuger, S., Klieme, E., Jude, N. & Kaplan, D. (Hrsg.), *Assessing contexts of learning: An international perspective*. Cham: Springer. doi:10.1007/978-3-319-45357-6_1.

ITC = International Test Commission (2005) International Test Commission guidelines for translating and adapting tests. http://www.intestcom.org/files/guideline_test_adaptation.pdf. [Accessed 01 September 2021].

Lenz, P., Aeppli, T., Karges, K., Wiedenkeller, E., Angelone, D. & Keller, F. (2019): Nationale Bildungsstandards in der ersten Fremdsprache. In Konsortium ÜGK (Hrsg.) (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2017: Sprachen 8. Schuljahr* (S. 35-49). Bern und Genf: EDK und SRED. https://doi.org/10.18747/PHSG-coll3/id/385

Linneweber-Lammerskitten, H., Angelone, D. und Keller, F. (2019). Überprüfte Kompetenzen in Mathematik. In Konsortium ÜGK (Hrsg.) (2019). *Überprüfung der Grundkompetenzen. Nationaler Bericht der ÜGK 2016: Mathematik 11. Schuljahr* (S. 17-27). Bern und Genf: EDK und SRED, https://doi.org/10.18747/PHSG

Maag Merki, K. (2016). Theoretische und empirische Analysen der Effektivität von Bildungsstandards, standardbezogenen Lernstandserhebungen und zentralen Abschlussprüfungen. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem*, 2., erweiterte Ausgabe (S. 151-182). VS Verlag für Sozialwissenschaften.

Pant, H. A. & Stanat, P. (2015). Wie kommt man zu guten Testaufgaben? Die Konstruktion von Aufgaben und Tests für die Ländervergleiche des IQB. *SEMINAR: Aufgaben. Im Leistungskontext. In Schule und Lehrerausbildung. 48. Seminartag Berlin (II), 21*(1), 5–18.

Pham, G., Helbling, L., Verner, M., Petrucci, F., Angelone, D. & Ambrosetti, A. (2019a). *ÜGK– COFO – VeCoF 2016 results: Technical appendices.* St.Gallen & Genf: Pädagogische Hochschule St.Gallen (PHSG) & Service de la recherche en éducation (SRED).

Pham, G., Helbling, L., Verner, M., & Ambrosetti, A. (2019b). *ÜGK – COFO – VeCoF 2017 results: Technical appendices.* St.Gallen & Genève: Pädagogische Hochschule St.Gallen (PHSG) & Service de la recherche en éducation (SRED).

Ramseier, E., Moser, U., Moreau, J. & Antonietti, J.-P. (2008). *Schlussbericht der HarmoS-Methodologiegruppe. Bern*. http://www.lehrplanforschung.ch/wp-content/uploads/2011/06/Methodologie-Schlussbericht-_HarmoS11.pdf

Rost, J. (2004): Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. *Zeitschrift für Pädagogik 50(5)*, S. 662-678.

Salvisberg, M. (2019). *Verifica delle competenze fondamentali nazionali 2017. Sviluppo degli esercizi nella lingua di scolarizzazione nell'8° anno scolastico.* Locarno: Scuola universitaria professionale delle Svizzera italiana (SUPSI).

Schneider, G., Lenz, P. & Studer, T. [Konsortium HarmoS Fremdsprachen] (2009). *Fremdsprachen: wissenschaftlicher Kurzbericht und Kompetenzmodell.* Bern: EDK. http://edudoc.ch/record/87025/files/L2_wissB_25_1_10_d.pdf

Tiffin-Richards, S.P. & Pant, H.A. (2017) Arguing Validity in Educational Assessment. In: D. Leutner, J. Fleischer, J. Grünkorn & E. Klieme (eds). *Competence Assessment in Education. Methodology of Educational Measurement and Assessment*. Springer, Cham. https://doi.org/10.1007/978-3-319-50030-0_27.

Verner, M., & Helbling, L. (2019). *Sampling ÜGK 2017. Technischer Bericht zu Stichprobendesign, Gewichtung und Varianzschätzung bei der Überprüfung des Erreichens der Grundkompetenzen 2017*. Zürich: Institut für Bildungsevaluation, assoziiertes Institut der Universität Zürich.

II. Anhang – Ergebnisse der standardisierten Befragung

Angaben zu Ihrem Arbeitsbereich

1. Was ist Ihr aktueller Arbeitsbereich? (mehrere Antworten möglich)

	Alle Befragten	Kantonale Vertreter*innen	Personen aus Forschung und Aus- und Weiterbildung
	Häufigkeit	Häufigkeit	Häufigkeit
a) Forschung	64 (45.4%)	1 (2.6%)	64 (83.1%)
b) Aus- und Weiterbildung von Lehrpersonen	44 (31.2%)	1 (2.6%)	44 (57.1%)
c) Schulpraxis	5 (3.5%)	0 (0%)	2 (2.6%)
d) Bildungsadministration: Leitende Position	24 (17%)	24 (61.5%)	2 (2.6%)
e) Bildungsadministration: Umsetzung der ÜGK	20 (14.2%)	20 (51.3%)	1 (1.3%)
f) Andere	10 (7.1%)	0 (0%)	2 (2.6%)
Subtotal	125	39	77
Fehlend	16 (11.3%)	0 (0%)	0 (0%)
Total	141	39	77

1a. Ist Ihr Kanton dem HarmoS-Konkordat beigetreten? (nur für Personen, die bei Frage 1 angegeben haben: d oder e

	Kantonale Vertreter*innen	
	Häufigkeit	Prozent
Ja	21	53.8%
Nein	15	38.5%
Noch offen	2	5.1%
Subtotal	38	97.4%
Fehlend	1	2.6%
Total	39	100.0%

1b. Welcher Disziplin bzw. welchen Disziplinen ordnen Sie Ihr Forschungsgebiet zu? (mehrere Antworten möglich)

	Personen aus Forschung und Aus- und Weiterbildung
	Häufigkeit
a) Erziehungswissenschaft	42 (54.5%)
b) Fachdidaktik(en)	21 (27.3%)
c) Psychologie	7 (9.1%)
d) Soziologie	15 (19.5%)
e) Politikwissenschaft	4 (5.2%)
f) (Bildungs-)Ökonomie	3 (3.9%)
g) Andere	4 (5.2%)
Subtotal	64
Fehlend	13 (16.9%)
Total	77

2. In welcher Sprachregion arbeiten Sie? (mehrere Antworten möglich)

	Alle Befragten	Kantonale Vertreter*innen	Personen aus Forschung und Aus- und Weiterbildung
	Häufigkeit	Häufigkeit	Häufigkeit
a) Romandie	24 (17%)	9 (23.1%)	14 (18.2%)
b) Italienische Schweiz	11 (7.8%)	2 (5.1%)	9 (11.7%)
c) Deutschschweiz	92 (65.2%)	28 (71.8%)	57 (74%)
Subtotal	123	39	76
Fehlend	18 (12.8%)	0 (0%)	1 (1.3%)
Total Antwortende	141	39	77

A. Grundsätzliche Bedeutung der ÜGK für die Qualitätsentwicklung des Bildungswesens

*3. Als wie wichtig erachten Sie es, dass das Erreichen der Grundkompetenzen der Schüler*innen auf nationaler Ebene überprüft wird?*

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Überhaupt nicht wichtig	2	1.4%	1	2.6%	2	2.6%
Eher nicht wichtig	10	7.1%	0	0.0%	10	13.0%
Eher wichtig	46	32.6%	17	43.6%	26	33.8%
Sehr wichtig	54	38.3%	20	51.3%	31	40.3%
Subtotal	112	79.4%	38	97.4%	69	89.6%
Fehlend	29	20.6%	1	2.6%	8	10.4%
Total	141	100.0%	39	100.0%	77	100.0%

4. Unabhängig von der aktuellen Durchführung: Wie wichtig ist die ÜGK aus Ihrer Sicht für das Fällen von bildungspolitischen Entscheiden:
a) auf nationaler Ebene?
b) auf sprachregionaler Ebene?
c) auf kantonaler Ebene?

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Überhaupt nicht wichtig	4	2.8%	1	2.6%	4	5.2%
	Eher nicht wichtig	20	14.2%	2	5.1%	17	22.1%
	Eher wichtig	38	27.0%	20	51.3%	16	20.8%
	Sehr wichtig	47	33.3%	15	38.5%	29	37.7%
	Subtotal	109	77.3%	38	97.4%	66	85.7%
	Fehlend	32	22.7%	1	2.6%	11	14.3%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Überhaupt nicht wichtig	6	4.3%	1	2.6%	6	7.8%
	Eher nicht wichtig	26	18.4%	12	30.8%	14	18.2%
	Eher wichtig	45	31.9%	17	43.6%	24	31.2%
	Sehr wichtig	31	22.0%	8	20.5%	21	27.3%
	Subtotal	108	76.6%	38	97.4%	65	84.4%
	Fehlend	33	23.4%	1	2.6%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Überhaupt nicht wichtig	6	4.3%	2	5.1%	5	6.5%
	Eher nicht wichtig	20	14.2%	6	15.4%	13	16.9%
	Eher wichtig	41	29.1%	14	35.9%	26	33.8%
	Sehr wichtig	42	29.8%	16	41.0%	22	28.6%
	Subtotal	109	77.3%	38	97.4%	66	85.7%
	Fehlend	32	22.7%	1	2.6%	11	14.3%
	Total	141	100.0%	39	100.0%	77	100.0%

5. Unabhängig von der aktuellen Durchführung: Wie wichtig ist die ÜGK aus Ihrer Sicht für das Fällen von Entscheiden für die Schulpraxis?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Überhaupt nicht wichtig	11	7.8%	3	7.7%	8	10.4%
Eher nicht wichtig	33	23.4%	11	28.2%	20	26.0%
Eher wichtig	42	29.8%	19	48.7%	20	26.0%
Sehr wichtig	24	17.0%	5	12.8%	19	24.7%
Subtotal	110	78.0%	38	97.4%	67	87.0%
Fehlend	31	22.0%	1	2.6%	10	13.0%
Total	141	100.0%	39	100.0%	77	100.0%

*6. Die ÜGK umfasst zwei Elemente: a) Leistungsmessungen in einzelnen Fächern und b) Kontextfragebögen für die Erfassung von Merkmalen der Schüler*innen und der Lernumgebung (z.B. soziale Herkunft, Motivation, Unterricht). Wie wichtig ist für Ihren Aufgaben- und Zuständigkeitsbereich...*
*a) die Erfassung der Leistungen der Schüler*innen?*
*b) die Erfassung von individuellen Merkmalen der Schüler*innen (z.B. soziale Herkunft)?*
c) die Erfassung von Merkmalen der Lernumgebung (z.B. Unterricht)?

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Überhaupt nicht wichtig	7	5.0%	0	0.0%	7	9.1%
	Eher nicht wichtig	11	7.8%	0	0.0%	11	14.3%
	Eher wichtig	35	24.8%	15	38.5%	16	20.8%
	Sehr wichtig	56	39.7%	22	56.4%	33	42.9%
	Subtotal	109	77.3%	37	94.9%	67	87.0%
	Fehlend	32	22.7%	2	5.1%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Überhaupt nicht wichtig	3	2.1%	0	0.0%	3	3.9%
	Eher nicht wichtig	9	6.4%	1	2.6%	7	9.1%
	Eher wichtig	41	29.1%	14	35.9%	25	32.5%
	Sehr wichtig	56	39.7%	22	56.4%	32	41.6%
	Subtotal	109	77.3%	37	94.9%	67	87.0%
	Fehlend	32	22.7%	2	5.1%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Überhaupt nicht wichtig	1	0.7%	0	0.0%	1	1.3%
	Eher nicht wichtig	6	4.3%	3	7.7%	3	3.9%
	Eher wichtig	42	29.8%	17	43.6%	22	28.6%
	Sehr wichtig	60	42.6%	17	43.6%	41	53.2%
	Subtotal	109	77.3%	37	94.9%	67	87.0%
	Fehlend	32	22.7%	2	5.1%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%

7. Aktuell hat die ÜGK die Funktion der Systemevaluation auf nationaler Ebene. Welche Funktion sollte aus Ihrer Sicht die ÜGK haben?
a) Systemevaluation auf nationaler Ebene
b) Systemevaluation auf sprachregionaler Ebene
c) Systemevaluation auf kantonaler Ebene
d) Identifikation von Ursachen allfälliger Unterschiede zwischen regionalen oder kantonalen Bildungssystemen
e) Handlungswissen, wie das Bildungssystem auf nationaler Ebene verändert werden kann
f) Handlungswissen, wie das Bildungssystem auf sprachregionaler Ebene verändert werden kann
g) Handlungswissen, wie das Bildungssystem auf kantonaler Ebene verändert werden kann
h) Handlungswissen, wie die Schulpraxis verändert werden kann
i) Andere

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Nein	0	0.0%	0	0.0%	0	0.0%
	Eher nein	7	5.0%	0	0.0%	7	9.1%
	Eher ja	25	17.7%	7	17.9%	17	22.1%
	Ja	77	54.6%	30	76.9%	43	55.8%
	Subtotal	109	77.3%	37	94.9%	67	87.0%
	Fehlend	32	22.7%	2	5.1%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Nein	4	2.8%	1	2.6%	3	3.9%
	Eher nein	19	13.5%	6	15.4%	13	16.9%
	Eher ja	33	23.4%	14	35.9%	16	20.8%
	Ja	54	38.3%	16	41.0%	36	46.8%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%

c)	Nein	7	5.0%	1	2.6%	6	7.8%
	Eher nein	19	13.5%	7	17.9%	10	13.0%
	Eher ja	34	24.1%	11	28.2%	22	28.6%
	Ja	50	35.5%	18	46.2%	30	39.0%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Nein	4	2.8%	1	2.6%	3	3.9%
	Eher nein	21	14.9%	5	12.8%	13	16.9%
	Eher ja	38	27.0%	20	51.3%	18	23.4%
	Ja	47	33.3%	11	28.2%	34	44.2%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
e)	Nein	8	5.7%	0	0.0%	8	10.4%
	Eher nein	17	12.1%	4	10.3%	12	15.6%
	Eher ja	36	25.5%	18	46.2%	15	19.5%
	Ja	49	34.8%	15	38.5%	33	42.9%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
f)	Nein	11	7.8%	1	2.6%	10	13.0%
	Eher nein	27	19.1%	12	30.8%	14	18.2%
	Eher ja	40	28.4%	18	46.2%	17	22.1%
	Ja	31	22.0%	6	15.4%	26	33.8%
	Subtotal	109	77.3%	37	94.9%	67	87.0%
	Fehlend	32	22.7%	2	5.1%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
g)	Nein	15	10.6%	2	5.1%	13	16.9%
	Eher nein	16	11.3%	6	15.4%	8	10.4%
	Eher ja	32	22.7%	15	38.5%	16	20.8%
	Ja	47	33.3%	14	35.9%	31	40.3%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
h)	Nein	20	14.2%	3	7.7%	16	20.8%
	Eher nein	28	19.9%	13	33.3%	12	15.6%
	Eher ja	27	19.1%	15	38.5%	11	14.3%
	Ja	35	24.8%	6	15.4%	29	37.7%
	Subtotal	110	78.0%	37	94.9%	68	88.3%
	Fehlend	31	22.0%	2	5.1%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
i)	Nein	16	11.3%	6	15.4%	7	9.1%
	Eher nein	4	2.8%	1	2.6%	3	3.9%
	Eher ja	3	2.1%	2	5.1%	1	1.3%
	Ja	4	2.8%	1	2.6%	3	3.9%
	Subtotal	27	19.1%	10	25.6%	14	18.2%
	Fehlend	114	80.9%	29	74.4%	63	81.8%
	Total	141	100.0%	39	100.0%	77	100.0%

8. Welchen Mehrwert, alles in allem, erbringt die ÜGK für Ihren Aufgaben- und Zuständigkeitsbereich im Vergleich zu anderen Daten und Informationen im Bildungssystem?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Überhaupt keinen Mehrwert	8	5.7%	1	2.6%	7	9.1%
Einen eher kleinen Mehrwert	47	33.3%	15	38.5%	29	37.7%
Einen eher grossen Mehrwert	39	27.7%	17	43.6%	21	27.3%
Einen sehr grossen Mehrwert	16	11.3%	5	12.8%	11	14.3%
Subtotal	110	78.0%	38	97.4%	68	88.3%
Fehlend	31	22.0%	1	2.6%	9	11.7%
Total	141	100.0%	39	100.0%	77	100.0%

B. Nutzung der ÜGK für den eigenen Arbeits- und Zuständigkeitsbereich

9. Wie häufig nutzen Sie in Ihrem Arbeits- und Zuständigkeitsbereich die nachfolgenden Informationen für die Identifikation von Stärken und Schwächen des Bildungssystems?
a) Ergebnisse aus den nationalen Berichten zur ÜGK
b) Vertiefende Analysen zur ÜGK
c) Ergebnisse aus Erhebungen / Tests im eigenen Kanton
d) Ergebnisse aus Erhebungen / Tests in der eigenen Sprachregion
e) Wissenschaftliche Publikationen im Allgemeinen
*f) Informationen aus Vernehmlassungen oder Beratungen im Kanton oder in der Sprachregion (z.B. 'Runder Tisch', regelmässige Treffen mit Akteur*innen aus der Schulpraxis)*
g) Bildungsberichte Schweiz
h) Andere

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Nie oder fast nie	25	17.7%	4	10.3%	20	26.0%
	Selten	48	34.0%	20	51.3%	25	32.5%
	Eher häufig	24	17.0%	11	28.2%	15	19.5%
	Sehr häufig	10	7.1%	1	2.6%	7	9.1%
	Subtotal	107	75.9%	36	92.3%	67	87.0%
	Fehlend	34	24.1%	3	7.7%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Nie oder fast nie	38	27.0%	7	17.9%	28	36.4%
	Selten	33	23.4%	14	35.9%	18	23.4%
	Eher häufig	29	20.6%	13	33.3%	17	22.1%
	Sehr häufig	8	5.7%	2	5.1%	5	6.5%
	Subtotal	108	76.6%	36	92.3%	68	88.3%
	Fehlend	33	23.4%	3	7.7%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Nie oder fast nie	31	22.0%	3	7.7%	28	36.4%
	Selten	34	24.1%	9	23.1%	22	28.6%
	Eher häufig	27	19.1%	13	33.3%	12	15.6%
	Sehr häufig	14	9.9%	10	25.6%	5	6.5%
	Subtotal	106	75.2%	35	89.7%	67	87.0%
	Fehlend	35	24.8%	4	10.3%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Nie oder fast nie	32	22.7%	6	15.4%	25	32.5%
	Selten	48	34.0%	16	41.0%	29	37.7%
	Eher häufig	21	14.9%	11	28.2%	10	13.0%
	Sehr häufig	7	5.0%	3	7.7%	4	5.2%
	Subtotal	108	76.6%	36	92.3%	68	88.3%
	Fehlend	33	23.4%	3	7.7%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
e)	Nie oder fast nie	8	5.7%	1	2.6%	7	9.1%
	Selten	11	7.8%	5	12.8%	4	5.2%
	Eher häufig	35	24.8%	17	43.6%	15	19.5%
	Sehr häufig	54	38.3%	13	33.3%	42	54.5%
	Subtotal	108	76.6%	36	92.3%	68	88.3%
	Fehlend	33	23.4%	3	7.7%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
f)	Nie oder fast nie	25	17.7%	1	2.6%	22	28.6%
	Selten	31	22.0%	7	17.9%	24	31.2%
	Eher häufig	33	23.4%	15	38.5%	16	20.8%
	Sehr häufig	17	12.1%	12	30.8%	5	6.5%
	Subtotal	106	75.2%	35	89.7%	67	87.0%
	Fehlend	35	24.8%	4	10.3%	10	13.0%
	Total	141	100.0%	39	100.0%	77	100.0%

g)	Nie oder fast nie	5	3.5%	0	0.0%	5	6.5%
	Selten	19	13.5%	7	17.9%	10	13.0%
	Eher häufig	58	41.1%	21	53.8%	36	46.8%
	Sehr häufig	26	18.4%	8	20.5%	17	22.1%
	Subtotal	108	76.6%	36	92.3%	68	88.3%
	Fehlend	33	23.4%	3	7.7%	9	11.7%
	Total	141	100.0%	39	100.0%	77	100.0%
h)	Nie oder fast nie	7	5.0%	0	0.0%	6	7.8%
	Selten	13	9.2%	6	15.4%	6	7.8%
	Eher häufig	6	4.3%	2	5.1%	4	5.2%
	Sehr häufig	7	5.0%	0	0.0%	5	6.5%
	Subtotal	33	23.4%	8	20.5%	21	27.3%
	Fehlend	108	76.6%	31	79.5%	56	72.7%
	Total	141	100.0%	39	100.0%	77	100.0%

10. Wie häufig nutzen Sie in Ihrem Arbeits- und Zuständigkeitsbereich die nachfolgenden Informationen für Entscheide bezüglich der Qualitätsentwicklung des Bildungswesens?
a) Ergebnisse aus den nationalen Berichten zur ÜGK
b) Vertiefende Analysen zur ÜGK
c) Ergebnisse aus Erhebungen / Tests im eigenen Kanton
d) Ergebnisse aus Erhebungen / Tests in der eigenen Sprachregion
e) Wissenschaftliche Publikationen im Allgemeinen
*f) Informationen aus Vernehmlassungen oder Beratungen im Kanton oder in der Sprachregion (z.B. 'Runder Tisch', regelmässige Treffen mit Akteur*innen aus der Schulpraxis)*
g) Bildungsberichte Schweiz
h) Andere

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Nie oder fast nie	31	22.0%	4	10.3%	25	32.5%
	Selten	45	31.9%	19	48.7%	24	31.2%
	Eher häufig	20	14.2%	9	23.1%	12	15.6%
	Sehr häufig	8	5.7%	4	10.3%	4	5.2%
	Subtotal	104	73.8%	36	92.3%	65	84.4%
	Fehlend	37	26.2%	3	7.7%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Nie oder fast nie	37	26.2%	8	20.5%	27	35.1%
	Selten	39	27.7%	14	35.9%	24	31.2%
	Eher häufig	20	14.2%	10	25.6%	10	13.0%
	Sehr häufig	7	5.0%	4	10.3%	4	5.2%
	Subtotal	103	73.0%	36	92.3%	65	84.4%
	Fehlend	38	27.0%	3	7.7%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Nie oder fast nie	24	17.0%	3	7.7%	21	27.3%
	Selten	33	23.4%	6	15.4%	25	32.5%
	Eher häufig	32	22.7%	15	38.5%	16	20.8%
	Sehr häufig	14	9.9%	12	30.8%	3	3.9%
	Subtotal	103	73.0%	36	92.3%	65	84.4%
	Fehlend	38	27.0%	3	7.7%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Nie oder fast nie	26	18.4%	5	12.8%	21	27.3%
	Selten	47	33.3%	16	41.0%	28	36.4%
	Eher häufig	25	17.7%	11	28.2%	15	19.5%
	Sehr häufig	5	3.5%	4	10.3%	1	1.3%
	Subtotal	103	73.0%	36	92.3%	65	84.4%
	Fehlend	38	27.0%	3	7.7%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%

e)	Nie oder fast nie	13	9.2%	0	0.0%	13	16.9%
	Selten	14	9.9%	6	15.4%	6	7.8%
	Eher häufig	34	24.1%	18	46.2%	16	20.8%
	Sehr häufig	41	29.1%	11	28.2%	30	39.0%
	Subtotal	102	72.3%	35	89.7%	65	84.4%
	Fehlend	39	27.7%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
f)	Nie oder fast nie	23	16.3%	1	2.6%	21	27.3%
	Selten	27	19.1%	5	12.8%	23	29.9%
	Eher häufig	30	21.3%	16	41.0%	12	15.6%
	Sehr häufig	22	15.6%	14	35.9%	8	10.4%
	Subtotal	102	72.3%	36	92.3%	64	83.1%
	Fehlend	39	27.7%	3	7.7%	13	16.9%
	Total	141	100.0%	39	100.0%	77	100.0%
g)	Nie oder fast nie	14	9.9%	0	0.0%	13	16.9%
	Selten	25	17.7%	10	25.6%	14	18.2%
	Eher häufig	48	34.0%	18	46.2%	31	40.3%
	Sehr häufig	15	10.6%	7	17.9%	7	9.1%
	Subtotal	102	72.3%	35	89.7%	65	84.4%
	Fehlend	39	27.7%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
h)	Nie oder fast nie	11	7.8%	0	0.0%	10	13.0%
	Selten	14	9.9%	6	15.4%	7	9.1%
	Eher häufig	4	2.8%	2	5.1%	2	2.6%
	Sehr häufig	3	2.1%	0	0.0%	3	3.9%
	Subtotal	32	22.7%	8	20.5%	22	28.6%
	Fehlend	109	77.3%	31	79.5%	55	71.4%
	Total	141	100.0%	39	100.0%	77	100.0%

11. Können Sie sich Änderungen im Konzept oder in der Vorgehensweise der ÜGK vorstellen, die dazu führen würden, dass Sie die Ergebnisse in Ihrem Aufgaben- und Zuständigkeitsbereich besser nutzen könnten?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Nein	42	29.8%	9	23.1%	31	40.3%
Ja	61	43.3%	26	66.7%	34	44.2%
Subtotal	103	73.0%	35	89.7%	65	84.4%
Fehlend	38	27.0%	4	10.3%	12	15.6%
Total	141	100.0%	39	100.0%	77	100.0%

C. Kommunikation zwischen verschiedenen Akteur*innen

12. Wie gut sind Sie über die ÜGK informiert?
a) über die Ziele der ÜGK
b) darüber, wer insgesamt strategisch für die ÜGK verantwortlich ist
c) über die involvierten Institutionen, die den Auftrag haben, die ÜGK durchzuführen
d) über die Art und Weise der Durchführung der Erhebungen (z.B. Rhythmus, Auswahl Fächer)
e) über die Ergebnisse der ÜGK

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Überhaupt nicht gut	6	4.3%	0	0.0%	6	7.8%
	Eher nicht gut	17	12.1%	3	7.7%	14	18.2%
	Eher gut	37	26.2%	11	28.2%	25	32.5%
	Sehr gut	45	31.9%	21	53.8%	20	26.0%
	Subtotal	105	74.5%	35	89.7%	65	84.4%
	Fehlend	36	25.5%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Überhaupt nicht gut	11	7.8%	0	0.0%	11	14.3%
	Eher nicht gut	21	14.9%	6	15.4%	16	20.8%
	Eher gut	34	24.1%	13	33.3%	19	24.7%
	Sehr gut	39	27.7%	16	41.0%	19	24.7%
	Subtotal	105	74.5%	35	89.7%	65	84.4%
	Fehlend	36	25.5%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Überhaupt nicht gut	12	8.5%	2	5.1%	10	13.0%
	Eher nicht gut	26	18.4%	6	15.4%	18	23.4%
	Eher gut	28	19.9%	12	30.8%	16	20.8%
	Sehr gut	39	27.7%	15	38.5%	21	27.3%
	Subtotal	105	74.5%	35	89.7%	65	84.4%
	Fehlend	36	25.5%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Überhaupt nicht gut	12	8.5%	0	0.0%	12	15.6%
	Eher nicht gut	24	17.0%	5	12.8%	18	23.4%
	Eher gut	38	27.0%	17	43.6%	20	26.0%
	Sehr gut	31	22.0%	13	33.3%	15	19.5%
	Subtotal	105	74.5%	35	89.7%	65	84.4%
	Fehlend	36	25.5%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%
e)	Überhaupt nicht gut	9	6.4%	0	0.0%	9	11.7%
	Eher nicht gut	23	16.3%	1	2.6%	20	26.0%
	Eher gut	35	24.8%	16	41.0%	20	26.0%
	Sehr gut	38	27.0%	18	46.2%	16	20.8%
	Subtotal	105	74.5%	35	89.7%	65	84.4%
	Fehlend	36	25.5%	4	10.3%	12	15.6%
	Total	141	100.0%	39	100.0%	77	100.0%

13. Mit welchen der nachfolgenden Berichte zur ÜGK haben Sie sich in welcher Tiefe auseinandergesetzt?
a) Mit dem nationalen Bericht zur ÜGK 2016, Mathematik 11. Schuljahr
b) Mit einzelnen oder allen weiterführenden Berichten zur ÜGK 2016 (z.B. zur Testentwicklung oder zum Kontextfragebogen)
c) Mit dem nationalen Bericht zur ÜGK 2017, Sprachen 8. Schuljahr
d) Mit einzelnen oder allen weiterführenden Berichten zur ÜGK 2017 (z.B. zur Testentwicklung oder zum Kontextfragebogen)
e) Mit zusätzlichen wissenschaftlichen Publikationen, die mit Daten der ÜGK entstanden sind

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Nicht	29	20.6%	2	5.1%	25	32.5%
	Oberflächlich	12	8.5%	1	2.6%	11	14.3%
	Eher oberflächlich	8	5.7%	3	7.7%	3	3.9%
	Eher vertieft	25	17.7%	15	38.5%	10	13.0%
	Vertieft	30	21.3%	14	35.9%	15	19.5%
	Subtotal	104	73.8%	35	89.7%	64	83.1%
	Fehlend	37	26.2%	4	10.3%	13	16.9%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Nicht	28	19.9%	3	7.7%	23	29.9%
	Oberflächlich	14	9.9%	3	7.7%	10	13.0%
	Eher oberflächlich	22	15.6%	13	33.3%	10	13.0%
	Eher vertieft	23	16.3%	11	28.2%	10	13.0%
	Vertieft	17	12.1%	5	12.8%	11	14.3%
	Subtotal	104	73.8%	35	89.7%	64	83.1%
	Fehlend	37	26.2%	4	10.3%	13	16.9%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Nicht	23	16.3%	1	2.6%	20	26.0%
	Oberflächlich	10	7.1%	1	2.6%	9	11.7%
	Eher oberflächlich	16	11.3%	3	7.7%	11	14.3%
	Eher vertieft	24	17.0%	14	35.9%	10	13.0%
	Vertieft	30	21.3%	16	41.0%	13	16.9%
	Subtotal	103	73.0%	35	89.7%	63	81.8%
	Fehlend	38	27.0%	4	10.3%	14	18.2%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Nicht	25	17.7%	2	5.1%	22	28.6%
	Oberflächlich	20	14.2%	4	10.3%	13	16.9%
	Eher oberflächlich	21	14.9%	13	33.3%	9	11.7%
	Eher vertieft	22	15.6%	10	25.6%	11	14.3%
	Vertieft	16	11.3%	6	15.4%	9	11.7%
	Subtotal	104	73.8%	35	89.7%	64	83.1%
	Fehlend	37	26.2%	4	10.3%	13	16.9%
	Total	141	100.0%	39	100.0%	77	100.0%
e)	Nicht	28	19.9%	6	15.4%	19	24.7%
	Oberflächlich	17	12.1%	9	23.1%	8	10.4%
	Eher oberflächlich	24	17.0%	12	30.8%	13	16.9%
	Eher vertieft	20	14.2%	5	12.8%	13	16.9%
	Vertieft	13	9.2%	3	7.7%	10	13.0%
	Subtotal	102	72.3%	35	89.7%	63	81.8%
	Fehlend	39	27.7%	4	10.3%	14	18.2%
	Total	141	100.0%	39	100.0%	77	100.0%

14. Wie beurteilen Sie die Qualität der folgenden Informationen zu den Ergebnissen der ÜGK?
a) Nationaler Bericht Mathematik 11. Schuljahr
b) Nationaler Bericht Sprachen 8. Schuljahr
c) Medienkonferenz
d) Presse
e) soziale Medien
f) Website der EDK
g) Website des Kantons
h) Website der ÜGK Schweiz http://uegk-schweiz.ch

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Nicht genutzt / nicht vorhanden	32	22.7%	2	5.1%	28	36.4%
	Ungenügende Qualität	0	0.0%	0	0.0%	0	0.0%
	Genügende Qualität	16	11.3%	8	20.5%	8	10.4%
	Gute Qualität	41	29.1%	21	53.8%	19	24.7%
	Sehr gute Qualität	10	7.1%	4	10.3%	4	5.2%
	Subtotal	99	70.2%	35	89.7%	59	76.6%
	Fehlend	42	29.8%	4	10.3%	18	23.4%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Nicht genutzt / nicht vorhanden	28	19.9%	1	2.6%	25	32.5%
	Ungenügende Qualität	0	0.0%	0	0.0%	0	0.0%
	Genügende Qualität	14	9.9%	8	20.5%	6	7.8%
	Gute Qualität	44	31.2%	21	53.8%	22	28.6%
	Sehr gute Qualität	11	7.8%	4	10.3%	5	6.5%
	Subtotal	97	68.8%	34	87.2%	58	75.3%
	Fehlend	44	31.2%	5	12.8%	19	24.7%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Nicht genutzt / nicht vorhanden	51	36.2%	7	17.9%	42	54.5%
	Ungenügende Qualität	5	3.5%	1	2.6%	4	5.2%
	Genügende Qualität	18	12.8%	13	33.3%	6	7.8%
	Gute Qualität	19	13.5%	13	33.3%	5	6.5%
	Sehr gute Qualität	3	2.1%	0	0.0%	0	0.0%
	Subtotal	96	68.1%	34	87.2%	57	74.0%
	Fehlend	45	31.9%	5	12.8%	20	26.0%
	Total	141	100.0%	39	100.0%	77	100.0%
d)	Nicht genutzt / nicht vorhanden	38	27.0%	5	12.8%	32	41.6%
	Ungenügende Qualität	16	11.3%	4	10.3%	12	15.6%
	Genügende Qualität	27	19.1%	16	41.0%	11	14.3%
	Gute Qualität	12	8.5%	9	23.1%	2	2.6%
	Sehr gute Qualität	3	2.1%	0	0.0%	0	0.0%
	Subtotal	96	68.1%	34	87.2%	57	74.0%
	Fehlend	45	31.9%	5	12.8%	20	26.0%
	Total	141	100.0%	39	100.0%	77	100.0%
e)	Nicht genutzt / nicht vorhanden	68	48.2%	18	46.2%	45	58.4%
	Ungenügende Qualität	12	8.5%	7	17.9%	5	6.5%
	Genügende Qualität	10	7.1%	5	12.8%	6	7.8%
	Gute Qualität	3	2.1%	3	7.7%	0	0.0%
	Sehr gute Qualität	1	0.7%	0	0.0%	1	1.3%
	Subtotal	94	66.7%	33	84.6%	57	74.0%
	Fehlend	47	33.3%	6	15.4%	20	26.0%
	Total	141	100.0%	39	100.0%	77	100.0%
f)	Nicht genutzt / nicht vorhanden	34	24.1%	4	10.3%	27	35.1%
	Ungenügende Qualität	7	5.0%	3	7.7%	4	5.2%
	Genügende Qualität	19	13.5%	6	15.4%	11	14.3%
	Gute Qualität	34	24.1%	20	51.3%	15	19.5%
	Sehr gute Qualität	3	2.1%	0	0.0%	1	1.3%
	Subtotal	97	68.8%	33	84.6%	58	75.3%
	Fehlend	44	31.2%	6	15.4%	19	24.7%
	Total	141	100.0%	39	100.0%	77	100.0%

g)	Nicht genutzt / nicht vorhanden	57	40.4%	14	35.9%	39	50.6%
	Ungenügende Qualität	8	5.7%	3	7.7%	4	5.2%
	Genügende Qualität	15	10.6%	7	17.9%	8	10.4%
	Gute Qualität	12	8.5%	8	20.5%	5	6.5%
	Sehr gute Qualität	1	0.7%	0	0.0%	1	1.3%
	Subtotal	93	66.0%	32	82.1%	57	74.0%
	Fehlend	48	34.0%	7	17.9%	20	26.0%
	Total	141	100.0%	39	100.0%	77	100.0%
h)	Nicht genutzt / nicht vorhanden	34	24.1%	7	17.9%	23	29.9%
	Ungenügende Qualität	3	2.1%	2	5.1%	1	1.3%
	Genügende Qualität	23	16.3%	8	20.5%	15	19.5%
	Gute Qualität	33	23.4%	17	43.6%	16	20.8%
	Sehr gute Qualität	4	2.8%	0	0.0%	3	3.9%
	Subtotal	97	68.8%	34	87.2%	58	75.3%
	Fehlend	44	31.2%	5	12.8%	19	24.7%
	Total	141	100.0%	39	100.0%	77	100.0%

15. Müsste die Kommunikation der Ergebnisse durch die Behörden zukünftig verstärkt oder eher reduziert werden?
a) Kommunikation durch die EDK
b) Kommunikation durch die kantonalen Behörden

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	(Eher) verstärkt	63	44.7%	17	43.6%	44	57.1%
	Gerade richtig	33	23.4%	17	43.6%	14	18.2%
	(Eher) reduziert	3	2.1%	1	2.6%	2	2.6%
	Subtotal	99	70.2%	35	89.7%	60	77.9%
	Fehlend	42	29.8%	4	10.3%	17	22.1%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	(Eher) verstärkt	53	37.6%	10	25.6%	39	50.6%
	Gerade richtig	41	29.1%	23	59.0%	18	23.4%
	(Eher) reduziert	4	2.8%	2	5.1%	2	2.6%
	Subtotal	98	69.5%	35	89.7%	59	76.6%
	Fehlend	43	30.5%	4	10.3%	18	23.4%
	Total	141	100.0%	39	100.0%	77	100.0%

16. Wie stark sind Sie in die Diskussion der Ergebnisse der ÜGK in Bezug auf das Ableiten von bildungspolitischen Entscheiden involviert?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Ich bin nicht involviert	48	34.0%	1	2.6%	43	55.8%
Ich bin involviert, aber nur wenig	25	17.7%	10	25.6%	13	16.9%
Ich bin relativ stark involviert	18	12.8%	13	33.3%	6	7.8%
Ich bin sehr stark involviert	13	9.2%	11	28.2%	2	2.6%
Subtotal	104	73.8%	35	89.7%	64	83.1%
Fehlend	37	26.2%	4	10.3%	13	16.9%
Total	141	100.0%	39	100.0%	77	100.0%

17. Wie stark sind Sie in die Diskussion der Ergebnisse der ÜGK in Bezug auf das Ableiten von Entscheiden für die Schulpraxis involviert?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Ich bin nicht involviert	50	35.5%	3	7.7%	42	54.5%
Ich bin involviert, aber nur wenig	32	22.7%	16	41.0%	15	19.5%
Ich bin relativ stark involviert	16	11.3%	10	25.6%	6	7.8%
Ich bin sehr stark involviert	6	4.3%	6	15.4%	1	1.3%
Subtotal	104	73.8%	35	89.7%	64	83.1%
Fehlend	37	26.2%	4	10.3%	13	16.9%
Total	141	100.0%	39	100.0%	77	100.0%

18. Möchten Sie stärker in die Diskussion der Ergebnisse in Bezug auf das Ableiten von bildungspolitischen Entscheiden involviert sein?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Nein	54	38.3%	20	51.3%	30	39.0%
Ja	49	34.8%	14	35.9%	34	44.2%
Subtotal	103	73.0%	34	87.2%	64	83.1%
Fehlend	38	27.0%	5	12.8%	13	16.9%
Total	141	100.0%	39	100.0%	77	100.0%

19. Möchten Sie stärker in die Diskussion der Ergebnisse in Bezug auf das Ableiten von Entscheiden für die Schulpraxis involviert sein?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Nein	53	37.6%	18	46.2%	30	39.0%
Ja	50	35.5%	16	41.0%	34	44.2%
Subtotal	103	73.0%	34	87.2%	64	83.1%
Fehlend	38	27.0%	5	12.8%	13	16.9%
Total	141	100.0%	39	100.0%	77	100.0%

D. Konzept der ÜGK

20. Welche der nachfolgenden Fachbereiche sollten aus Ihrer Sicht im Rahmen der ÜGK überprüft werden? (mehrere Antworten möglich)

	Alle Befragten	Kantonale Vertreter*innen	Personen aus Forschung und Aus- und Weiterbildung
	Häufigkeit	Häufigkeit	Häufigkeit
Schulsprache	96 (68.1%)	35 (89.7%)	56 (72.7%)
Fremdsprachen	74 (52.5%)	29 (74.4%)	40 (51.9%)
Mathematik	95 (67.4%)	35 (89.7%)	55 (71.4%)
Natur, Mensch, Gesellschaft bzw. Naturwissenschaften	77 (54.6%)	23 (59%)	49 (63.6%)
Gestalten	20 (14.2%)	5 (12.8%)	14 (18.2%)
Musik	24 (17%)	8 (20.5%)	15 (19.5%)
Bewegung und Sport	23 (16.3%)	9 (23.1%)	13 (16.9%)
Subtotal	98	35	58
Fehlend	43 (30.5%)	4 (10.3%)	19 (24.7%)
Total	141	39	77

21. Als wie sinnvoll erachten Sie es, dass die Überprüfung des Erreichens der Grundkompetenzen auf das Ende des 4., 8. und 11. Schuljahres festgelegt ist?
a) Ende des 4. Schuljahres (ca. 8-Jährige)
b) Ende des 8. Schuljahres (ca. 12-Jährige)
c) Ende des 11. Schuljahres (ca. 15-Jährige)

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Überhaupt nicht sinnvoll	10	7.1%	2	5.1%	7	9.1%
	Eher nicht sinnvoll	18	12.8%	7	17.9%	12	15.6%
	Eher nicht sinnvoll	37	26.2%	13	33.3%	23	29.9%
	Sehr sinnvoll	34	24.1%	12	30.8%	18	23.4%
	Subtotal	99	70.2%	34	87.2%	60	77.9%
	Fehlend	42	29.8%	5	12.8%	17	22.1%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Überhaupt nicht sinnvoll	2	1.4%	0	0.0%	2	2.6%
	Eher nicht sinnvoll	7	5.0%	0	0.0%	5	6.5%
	Eher nicht sinnvoll	32	22.7%	10	25.6%	22	28.6%
	Sehr sinnvoll	58	41.1%	24	61.5%	31	40.3%
	Subtotal	99	70.2%	34	87.2%	60	77.9%
	Fehlend	42	29.8%	5	12.8%	17	22.1%
	Total	141	100.0%	39	100.0%	77	100.0%
c)	Überhaupt nicht sinnvoll	1	0.7%	0	0.0%	1	1.3%
	Eher nicht sinnvoll	3	2.1%	0	0.0%	3	3.9%
	Eher nicht sinnvoll	27	19.1%	9	23.1%	17	22.1%
	Sehr sinnvoll	68	48.2%	25	64.1%	39	50.6%
	Subtotal	99	70.2%	34	87.2%	60	77.9%
	Fehlend	42	29.8%	5	12.8%	17	22.1%
	Total	141	100.0%	39	100.0%	77	100.0%

22. Als wie sinnvoll erachten Sie es, dass Bildungsziele jeweils nur als Mindeststandard (Grundkompetenzen) beschrieben sind?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Überhaupt nicht sinnvoll	6	4.3%	2	5.1%	4	5.2%
Eher nicht sinnvoll	32	22.7%	12	30.8%	18	23.4%
Eher nicht sinnvoll	30	21.3%	10	25.6%	16	20.8%
Sehr sinnvoll	31	22.0%	10	25.6%	22	28.6%
Subtotal	99	70.2%	34	87.2%	60	77.9%
Fehlend	42	29.8%	5	12.8%	17	22.1%
Total	141	100.0%	39	100.0%	77	100.0%

23. Reicht es aus Ihrer Sicht, wenn die Tests einzig feststellen, ob die Grundkompetenzen erreicht werden, oder sollten die Tests das gesamte Leistungsspektrum mit mehreren Kompetenzstufen abdecken?
a) Anhand der Tests soll einzig überprüft werden, ob die Grundkompetenzen erreicht worden sind.
b) Die Tests sollen das gesamte Leistungsspektrum mit mehreren Kompetenzstufen abdecken (wie z.B. bei PISA).

		Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
a)	Stimme überhaupt nicht zu	21	14.9%	8	20.5%	13	16.9%
	Stimme eher nicht zu	28	19.9%	9	23.1%	19	24.7%
	Stimme eher zu	17	12.1%	5	12.8%	10	13.0%
	Stimme voll und ganz zu	30	21.3%	10	25.6%	17	22.1%
	Subtotal	96	68.1%	32	82.1%	59	76.6%
	Fehlend	45	31.9%	7	17.9%	18	23.4%
	Total	141	100.0%	39	100.0%	77	100.0%
b)	Stimme überhaupt nicht zu	16	11.3%	5	12.8%	10	13.0%
	Stimme eher nicht zu	19	13.5%	8	20.5%	9	11.7%
	Stimme eher zu	27	19.1%	6	15.4%	19	24.7%
	Stimme voll und ganz zu	37	26.2%	14	35.9%	23	29.9%
	Subtotal	99	70.2%	33	84.6%	61	79.2%
	Fehlend	42	29.8%	6	15.4%	16	20.8%
	Total	141	100.0%	39	100.0%	77	100.0%

24. Sollte mittelfristig an einer weitgehenden Umstellung der ÜGK auf ein digitales Assessment, das von den Lehrkräften selbst durchgeführt wird, gearbeitet werden?

	Alle Befragten		Kantonale Vertreter*innen		Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent	Häufigkeit	Prozent	Häufigkeit	Prozent
Nein	24	17.0%	6	15.4%	17	22.1%
Eher nein	25	17.7%	8	20.5%	14	18.2%
Eher ja	32	22.7%	11	28.2%	20	26.0%
Ja	14	9.9%	7	17.9%	8	10.4%
Subtotal	95	67.4%	32	82.1%	59	76.6%
Fehlend	46	32.6%	7	17.9%	18	23.4%
Total	141	100.0%	39	100.0%	77	100.0%

E. Einbindung der ÜGK in Wissenschaft und Forschung[23]

25. In der Schweiz wurde das Interfaculty Centre for Educational Research (ICER) an der Universität Bern mit der Koordination der ÜGK betraut. Wie sichtbar ist aus Ihrer Sicht das ICER im Kontext der Bildungsforschung in der Schweiz?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
Überhaupt nicht sichtbar	12	18.8%
Eher nicht sichtbar	26	40.6%
Eher sichtbar	10	15.6%
Sehr sichtbar	1	1.6%
Subtotal	49	76.6%
Fehlend	15	23.4%
Total	64	100.0%

26. Wie wichtig ist es Ihnen, im Zusammenhang mit Ihrer eigenen Forschung mit den wissenschaftlichen Institutionen der ÜGK (ICER, Aufgabendatenbank) in den Austausch zu treten?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
Überhaupt nicht wichtig	9	14.1%
Eher nicht wichtig	13	20.3%
Eher wichtig	15	23.4%
Sehr wichtig	13	20.3%
Subtotal	50	78.1%
Fehlend	14	21.9%
Total	64	100.0%

27. Soll dieser Austausch in Zukunft eher verstärkt oder reduziert werden, oder ist er gerade richtig?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
(Eher) verstärkt	24	37.5%
Gerade richtig	24	37.5%
(Eher) reduziert	1	1.6%
Subtotal	49	76.6%
Fehlend	15	23.4%
Total	64	100.0%

[23] Diese Fragen wurden nur den Personen gestellt, die als ihren aktuellen Arbeitsbereich ‘Forschung an Hochschulen oder Forschungsstellen’ angegeben haben: N = 64.

28. Sind Sie interessiert, an der Entwicklung der Leistungstests oder der Konzipierung des Kontextfragebogens mitzuarbeiten?
a) Leistungstests
b) Kontextfragebogen

		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent
a)	Überhaupt nicht interessiert	16	25.0%
	Eher nicht interessiert	15	23.4%
	Eher interessiert	8	12.5%
	Sehr interessiert	11	17.2%
	Subtotal	50	78.1%
	Fehlend	14	35.1%
	Total	64	100.0%
b)	Überhaupt nicht interessiert	12	18.8%
	Eher nicht interessiert	12	18.8%
	Eher interessiert	12	18.8%
	Sehr interessiert	13	20.3%
	Subtotal	49	76.6%
	Fehlend	15	23.4%
	Total	64	100.0%

29. Haben Sie die Berichterstattung zu den Ergebnissen der ÜGK als Ausgangspunkt für eigene Forschungsvorhaben genutzt?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
Nein	36	56.3%
Ja	11	17.2%
Ist geplant	3	4.7%
Subtotal	50	78.1%
Fehlend	14	21.9%
Total	64	100.0%

30 Haben Sie über FORS Daten der ÜGK für eigene Auswertungen bezogen?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
Nein	39	60.9%
Ja	11	17.2%
Subtotal	50	78.1%
Fehlend	14	21.9%
Total	64	100.0%

31. Planen Sie in Zukunft, über FORS Daten der ÜGK für eigene Auswertungen zu beziehen?

	Personen aus Forschung und Aus- und Weiterbildung	
	Häufigkeit	Prozent
Nein	16	25.0%
Eventuell	22	34.4%
Ja	11	17.2%
Subtotal	49	76.6%
Fehlend	15	23.4%
Total	64	100.0%

32. Können Sie sich Änderungen im Konzept oder in den Instrumenten (Tests, Kontextfragebogen) der ÜKG vorstellen, die dazu führen würden...

		Personen aus Forschung und Aus- und Weiterbildung	
		Häufigkeit	Prozent
a) dass mehr oder relevantere Fragen der Bildungsforschung bearbeitet werden können?	Nein	23	35.9%
	Ja	25	39.1%
	Subtotal	48	75.0%
	Fehlend	16	25.0%
	Total	64	100.0%
b) dass eine breitere Einbeziehung der Forscher*innen erreicht werden kann?	Nein	24	37.5%
	Ja	24	37.5%
	Subtotal	48	75.0%
	Fehlend	16	25.0%
	Total	64	100.0%
c) dass Sie die Daten der ÜGK für Ihre eigene Forschung häufiger nutzen können?	Nein	24	37.5%
	Ja	24	37.5%
	Subtotal	48	75.0%
	Fehlend	16	25.0%
	Total	64	100.0%

III. Anhang – Fragekatalog der EDK

Auszug aus:
Gesamtschau zur Überprüfung des Erreichens der Grundkompetenzen (ÜGK). Konzept. Fassung vom 12. Mai 2021

2. Hauptteil: Analyse

Die Grundlage für den zweiten Hauptteil der Gesamtschau bilden

- die Fragen an die Gutachter/-innen,
- die darauf basierende Analyse und die Antworten der Gutachter/-innen,
- Fachliche Stellungnahmen von weiteren externen Expert/-innen, insbesondere von Stefan Wolter, Direktor der Schweiz. Koordinationsstelle für Bildungsforschung SKBF,
- die Diskussionen des Gutachtens, der Stellungnahmen und allgemein von Fragen zur ÜGK im Kosta HarmoS.

Die Auftragsskizze an die Gutachter/-innen (Stand 12.05.2021) umfasst die im Folgenden abgedruckten Fragen, unterteilt nach den Bereichen externe und interne Governance. Sie decken im Wesentlichen die Fragen ab, die bei der Vorbereitung der Gesamtschau aufgeworfen wurden. Im Folgenden sind sie für die Gutachter/-innen kurz kontextualisiert.

2.1 Bereich externe Governance (Bereich A)

A1 Systemsteuerung, Zielkohärenz und Handlungskoordination

Hintergrund: Die bildungspolitischen Diskussionen der Nullerjahre führten zur Festlegung von Bildungszielen («Grundkompetenzen») und zur Konzipierung der ÜGK. Die ÜGK soll die Steuerung bestimmter Aspekte des Bildungssystems ermöglichen sowie insbesondere die gesamtschweizerische Harmonisierung der Bildungsziele über Kantone und Regionen hinweg überprüfen helfen.
Das Projekt ÜGK betrifft zahlreiche Akteur/-innen unterschiedlicher Ebenen: die EDK, die EDK-Regionalkonferenzen, die Kantone, die Schulen; die Geschäftsstelle Aufgabendatenbank (GS ADB) der EDK, das wissenschaftliche Konsortium Bildungsforschung und -evaluation; das Interfaculty Centre for Educational Research (ICER; Uni Bern) und dessen Projektpartner; die Bildungsforschung, die Berufsbildung und weitere. Im Zusammenhang mit dem Ansatz der ÜGK und deren Akteur/-innen stellen sich v.a. folgende Fragen:

Übergeordnete Frage:

A1.1 Ist der gewählte *Systemsteuerungsansatz* aufgrund des Kenntnisstandes des Jahres 2021 *grundsätzlich* noch aktuell? Wenn ja: Was kann dieser Steuerungsansatz konkret für die Qualitätsentwicklung des Bildungssystems leisten, was nicht? Wie kann seine Wirksamkeit gesteigert werden? Welche Mittel zur Systemsteuerung müssten ggf. ergänzt werden? Wenn nein: Was müsste an seine Stelle treten?

Weitere Fragen:

A1.2 Erkennen die Akteur/-innen in den *Zielen der ÜGK*[24] eine Relevanz – insgesamt und für den eigenen Arbeits- oder Zuständigkeitsbereich, die eigenen Aktivitäten? Welchen anderen/alternativen Zielen von Studien im schulischen Bereich räumen die Akteur/-innen ggf. Priorität ein?

A1.3 Inwiefern und aus welchen Gründen werden die Ergebnisse der ÜGK genutzt bzw. nicht genutzt?

A1.4 Wie gut gelingt die *Kommunikation* zwischen dem Projekt ÜGK und verschiedenen anderen Akteur/-innen, namentlich der Bildungspolitik und -administration auf unterschiedlichen Ebenen des

[24] aus dem Organisationsreglement ÜGK, Art. 2 Ziel der ÜGK:

1 Die ÜGK liefert den Kantonen auf der Ebene des Bildungssystems Informationen darüber, inwieweit die Grundkompetenzen auf bestimmten Schulstufen erreicht worden sind. Ziel der ÜGK ist die Evaluation von Leistungen des Bildungssystems der obligatorischen Schule und nicht die Beurteilung einzelner Schulen oder Lehrpersonen.

2 Die Ergebnisse der ÜGK fliessen in die gemeinsame Berichterstattung von Bund und Kantonen im Rahmen des Bildungsmonitorings ein (Artikel 61a BV).

föderalistischen Systems; der Bildungsforschung; den Schulen; der Öffentlichkeit? Wird das Wissen über die ÜGK, deren Funktion und Potential nachhaltig vermittelt? Gibt es genügend Austausch zwischen den Akteur/innen, damit die jeweiligen Anliegen gegenseitig wahrgenommen werden?

A1.5 Als wie gut und nützlich wird die Berichterstattung zu den Ergebnissen der ÜGK (ÜGKs 2016 und 2017) wahrgenommen?

A1.6 Werden Berichterstattung und Daten genutzt (auch für eigene weiterführende, z.B. lokale, Forschungsvorhaben)?

A1.7 Wie können ggf. *Zielkohärenz und Kommunikation* verbessert werden, um den Nutzen der ÜGK als Instrument zur Steuerung des Systems obligatorische Schule zu optimieren?

A2 Die ÜGK und weitere Erhebungen im Bildungssystem

Hintergrund: Die Schweiz beteiligt sich seit dem Jahr 2000 an den PISA-Studien (Kontextfragebögen für Schüler/-innen und Schulleitungen). Eine Beteiligung an PISA-Optionen wie *Financial Literacy* oder *Foreign Language Assessment* oder der *Study of Social and Emotional Skills* (SSES) sowie weiteren internationalen Studien im Bereich obligatorische Schule ist nicht grundsätzlich ausgeschlossen. Die ÜGK umfasst jeweils neben den kognitiven Tests auch einen Schülerfragebogen, der allerdings wegen der Belastung, v.a. der jüngeren Schüler/-innen, im Umfang recht beschränkt ist. Die Variablen und Konstrukte des Fragebogens dienen einerseits der Erklärung der Testergebnisse (Bsp. SES, Geschlecht) und sind andererseits (auch) für sich selbst von Interesse (Bsp. Einstellung zur Schule, IT-Literacy).

Als die ÜGK konzipiert wurde (vgl. Umsetzungsbeschluss der EDK zum HarmoS-Konkordat von 2007), sah man vor, in gesamtschweizersicher Absprache einerseits Referenztests zur Systemevaluation auf nationaler, regionaler und kantonaler Ebene einzuführen (Ziel: Bildungsmonitoring) und andererseits Tests zur individuellen Standortbestimmung (Ziel: individuelle Förderung). Die Instrumente für die Systemevaluation sollten auf nationaler Ebene erarbeitet werden, die Instrumente für die individuelle Standortbestimmung auf sprachregionaler Ebene. Zurzeit ist in den Regionen und Kantonen der Schweiz eine Reihe von externen Schulleistungstests (z.B. *Stellwerk, Checks*, kantonale Prüfungen, Orientierungsarbeiten, Abschlussprüfungen) im Einsatz. Die Kantonsumfrage von IDES (EDK) ermittelt jeweils, welche dieser Tests in den einzelnen Kantonen im Einsatz sind. Die ursprünglich vorgesehene Art von individuellen Standortbestimmungen hat sich nicht umfassend und gleichmässig durchgesetzt.

Im Hinblick darauf stellen sich insbesondere die folgenden Fragen:

A2.8 Welche Leistungen erbringen die Erhebungen und Erhebungsteile auf verschiedenen Ebenen des Bildungswesens (Individuum, Einzelschule, Kanton, Region, Schweiz) für die Systemsteuerung? Die bisherigen Konzeptionen des Zusammenwirkens der Ebenen stammen aus der Zeit vor 2007. Was sind neue Erkenntnisse dazu aufgrund des Forschungsstands von 2021? Welche Folgerungen ergeben sich für den künftigen Einsatz der verschiedenen Erhebungen?

A2.9 Wo bestehen Lücken bei der Beschaffung von Evidenz für die Steuerung des Bildungssystems, und mit welchen Massnahmen bzw. Erhebungen können diese Lücken gefüllt werden?

A2.10 Wie kann der Beitrag der verschiedenen Erhebungen zum schweizerischen Bildungsbericht optimiert werden? Dieser beleuchtet periodisch pro Bildungsstufe die Aspekte *Effectiveness, Efficiency* und *Equity* des Systems.

A2.11 Wie kann die ÜGK optimal zur Dateninfrastruktur beitragen, wie kann sie von der Dateninfrastruktur (auch von den Registerdaten des Bundesamtes für Statistik BFS) profitieren?

2.2 Bereich interne Governance (Bereich B)

Im Bereich der internen Governance sollen strukturell-organisatorische, fachliche und technische Aspekte der Umsetzung des ÜGK-Projekts fokussiert werden.

Der Auditbericht von Fischbach und Ugen (2018) zur ÜGK 2016 im Fachbereich Mathematik konzentrierte sich auf einen Teilbereich, nämlich auf fachliche Aspekte des Testens: Aufgaben- und Testentwicklung; Bezug zwischen der Beschreibung der Grundkompetenzen und dem Test; Prozess der Schwellenwertsetzung. Die Empfehlungen aufgrund der Befunde werden bei der Entwicklung der Aufgaben und Tests für das 4. Schuljahr heute bereits umgesetzt. Die Aufgabenentwicklung wird durch eine Ausweitung des Feedbackprozesses und eine Systematisierung, Standardisierung und ebenfalls eine quantitative

Vergrösserung der Prä-Pilotierung und Pilotierung deutlich verstärkt. Dazu kommen Arbeiten zur Konkretisierung der Beschreibung der Grundkompetenzen im Fachbereich Mathematik. Eine erneute Analyse der Aspekte, die im Auditbericht behandelt wurden, anhand der ÜGKs 2016 und 2017 erscheint deshalb nicht sinnvoll. Vielmehr sollte das Augenmerk auf das Vorgehen bei der Vorbereitung der ÜGK 2024 im 4. Schuljahr gerichtet werden.

B1 Erhebungsdesign

Hintergrund: Bisher hat die EDK für vier Fachbereiche (Schulsprache, Mathematik, Fremdsprachen, Naturwissenschaften) und jeweils das Ende der drei Schulstufen, welche die obligatorische Schule kennt, Bildungsziele in Form von Grundkompetenzen festgelegt (bei den Fremdsprachen nur für die zweite und dritte). Bei den Grundkompetenzen handelt es sich um Leistungsstandards in Form von Mindeststandards. Das Erreichen der Grundkompetenzen wird in grösseren Abständen stichprobenbasiert im Sinn eines schweizweiten Monitorings überprüft. Dabei sollen die Leistungen nicht auf einer Kompetenzskala verortet werden, sondern es wird überprüft, welcher Anteil der Schülerinnen und Schüler die Grundkompetenzen erreicht.

B1.1 Ist es sinnvoll, *vier Fachbereiche* (Schulsprache, Mathematik, Naturwissenschaften und Fremdsprachen), und gerade diese vier Fachbereiche für die Definition von gemeinsamen Bildungszielen und die Überprüfung des Erreichens der Bildungsziele auszuwählen?

B1.2 Ist es sinnvoll, für das Ende des *4., 8. und 11. Schuljahres* (nach HarmoS-Zählung, also für ca. 8-, 12- und 15-Jährige) Ziele zu beschreiben und deren Erreichung auf jeder dieser Stufen zu überprüfen?

B1.3 Ist es sinnvoll, in Kantonen, in denen keine Vollerhebung stattfindet, mit Schulstichproben und darin Schülerstichproben zu arbeiten anstatt mit Klassenstichproben?

B1.4 Ist es sinnvoll, als Bildungsziele jeweils *Mindeststandards* zu beschreiben und nur die Erreichung der Mindeststandards (ja/nein) zu überprüfen?

B1.5 Was ist ein optimaler *Erhebungsrhythmus* für jeden der Fachbereiche? Soll mit Längsschnitten gearbeitet werden (für die ÜGK 2020 im 11. Schuljahr war eine Stichprobe aus derselben Population vorgesehen, die bereits 2017 im 8. Schuljahr getestet wurde).

B1.6 Müssen die Erhebungen zur ÜGK in allen Kantonen zwingend in derselben Zeitperiode und jeweils in denselben Fächern durchgeführt werden? Sind flexiblere Modelle denkbar?

B2 Projektorganisation und institutionelle Verankerung

Hintergrund: Im Jahr 2018 wurde die Governance der ÜGK grundlegend verändert, sodass sich eine Begutachtung der früheren Strukturen (vgl. dazu v.a. das Konzept zur ÜGK von 2013) für sich allein genommen erübrigt. Das GS EDK zog sich aus der Umsetzungsebene zurück, indem es mit der Universität Bern eine umfassende, projektbezogene Kooperationsvereinbarung «betreffend die Zusammenarbeit zur Durchführung der ÜGK-Erhebungen in den Jahren 2020 und 2022 und zur kontinuierlichen wissenschaftlichen Koordination der Überprüfung des Erreichens der Grundkompetenzen (ÜGK)» abschloss. Das von der Universität Bern für die Umsetzung der Kooperation eingerichtete Interfaculty Centre for Educational Research (ICER) kooperiert seinerseits im Rahmen des Organisationsreglements für die Durchführung der ÜGK aus dem Jahr 2014 mit verschiedenen Institutionen eines wissenschaftlichen Konsortiums, um die ÜGK durchzuführen und wissenschaftlich einzubetten. Um Synergien nutzen zu können, wurde die Umsetzung und Koordination des PISA-Zyklus 2021 bzw. 2022 ebenfalls dem ICER übertragen. Dieses kooperiert auch für PISA mit dem wissenschaftlichen Konsortium und mit seinen Projektpartnern (Durchführungsebene). Die Koordination eines wissenschaftlichen Konsortiums zur Einbindung der Large-Scale-Assessments in die Wissenschaft sowie für die wissenschaftliche Vernetzung gehört zu den Aufgaben des ICER. Für die Umsetzung der Studien koordiniert das ICER die Projektleitung ÜGK bzw. PISA, die sich aus Vertreter/-innen der regionalen Durchführungszentren, des IBE (zuständig für die Stichprobenziehung), der PH GR (Datenmanagement) und der GS ADB (nur ÜGK) zusammensetzt.

Die GS ADB hat bei der Umsetzung der ÜGK eine Sonderstellung inne, indem sie strategisch und fachlich direkt dem GS EDK unterstellt ist. Zu ihren Aufgaben gehören insbesondere die Aufgaben- und Testentwicklung in Zusammenarbeit mit Fachdidaktiker/-innen und weiteren Expert/-innen, die Skalierung der Testergebnisse und die Schwellenwertsetzung. Die fachliche Anbindung an das GS EDK hat insbesondere damit zu tun, dass die ADB «Hüterin der Grundkompetenzen» ist. Das Organisationsreglement

für die ADB aus dem Jahr 2014 sieht im Zusammenhang mit Testaufgaben, Testskalen und Schwellenwerten auch weitergehende Dienstleistungen im Bildungssystem Schweiz vor (vgl. die Bezeichnung Aufgabendatenbank). Diese Funktionen sind allerdings gegenüber dem Beitrag zur ÜGK in den Hintergrund getreten und werden nicht ausgeübt. Organisatorisch und räumlich ist die GS ADB der Schweizerischen Koordinationsstelle für Bildungsforschung (SKBF) zugeordnet.

In Bezug auf die organisatorischen Aspekte stehen die folgenden Fragen im Vordergrund:

B2.7 Wie sind die Effektivität, Effizienz und Nachhaltigkeit der gewählten Organisationsstruktur im Hinblick auf die Durchführung der ÜGK einzuschätzen?

B2.8 Wie sind speziell die Positionierung und Anbindung der GS ADB einzuschätzen? Sollen Alternativen zur aktuellen Struktur gesucht werden? Welche wird ggf. empfohlen? Soll der ursprünglich (im Organisationsreglement) vorgesehene Funktionsumfang beibehalten und praktisch umgesetzt werden?

B2.9 Wie ist die Einbindung der Fachdidaktiken der getesteten Fachbereiche einzuschätzen, die nicht direkt Teil des wiss. Konsortiums und der Projektleitung ÜGK sind, sondern (bloss) bei der Aufgaben- und Testentwicklung mit der GS ADB zusammenarbeiten?

B2.10 Wie gut gelingen die Einbindung der ÜGK in Wissenschaft und Forschung sowie die Zusammenarbeit mit wissenschaftlichen Institutionen (vgl. Koordinationsauftrag des ICER)? Wie könnte die ÜGK von wissenschaftlichen Netzwerken und wissenschaftlicher Zusammenarbeit weiter profitieren?

B3 Fachliche und technische Aspekte

In Bezug auf die Projektumsetzung stehen verschiedene fachliche und technische Fragen im Raum.

B3.11 Bei der Testentwicklung für die ÜGK 2024 im 4.Schuljahr wurden aufgrund des Gutachtens von Fischbach & Uger (2018) Verbesserungen vorgenommen. Kann dieses Vorgehen als Muster für zukünftige ÜGK-Tests gelten?

B3.12 Die ADB leistet die Aufgaben- und Testentwicklung in Zusammenarbeit mit Fachdidaktiker/-innen aus den verschiedenen Sprachregionen (oft PH-Dozent/-innen). Anderswo leisten Personen mit anderem Hintergrund die Testentwicklung (z.B. Lehrpersonen). Gibt es für die ÜGK empfehlenswerte, glaubwürdige Alternativen zur aktuellen Praxis?

B3.13 Die Grundkompetenzen für die verschiedenen Fachbereiche sind auf unterschiedliche Art und Weise beschrieben, mehr oder weniger abstrakt, mit mehr oder weniger Bezug zu Aufgaben- und Leistungsmerkmalen. Welche Art der Beschreibung eignet sich besonders, um die erwarteten Grundkompetenzen allgemein verständlich und zugleich auf die Testskala beziehbar zu machen?

B3.14 Bei der Schwellenwertsetzung wird das Bookmarkverfahren eingesetzt. Verschiedentlich wurden die gesetzten Schwellen in Zweifel gezogen. Was wird empfohlen, um die Glaubwürdigkeit der Schwellenwertsetzungen zu erhalten bzw. zu erhöhen?

B3.15 Die ÜGK-Tests fokussieren auf den Schwellenwert, d.h. die Items zielen idealerweise auf Anforderungen, die noch im Bereich der Grundkompetenzen und leicht darüber liegen. Da in der ÜGK das Erreichen der Grundkompetenzen überprüft wird, werden die Kompetenzen von viel stärkeren Schüler/-innen nicht zuverlässig getestet (*ceiling effect*), was es beispielsweise erschwert, Aussagen zu Bezügen zwischen Testergebnissen und erklärenden Variablen zu machen. Inwiefern ist ein Festhalten am Testen von Grundkompetenzen angezeigt? Was sind ggf. sinnvolle Alternativen (vgl. Frage B 1.4)?

B3.16 Schülerfragebögen müssen oft in kurzer Erhebungszeit bei Schüler/-innen im Alter von ca. 8 bis 15 Jahren alle Informationen erfassen, die für die Kontextualisierung der Testergebnisse relevant sind und nicht anders besorgt werden können. Im Hinblick auf die Systemsteuerung und aufgrund von legitimen Interessen von Forschenden besteht ein Interesse an weiteren Variablen und Konstrukten. Zudem möchte man gleiche bzw. analoge Informationen über die ganze Altersspanne hinweg erheben können. Wie ist das Vorgehen einzuschätzen, das im ÜGK-Konsortium zurzeit verfolgt wird? Welche bewährten Alternativen gibt es?

B3.17 Bei den ÜGK-Erhebungen gehen Testadministrator/-innen mit Tablets in die Schulen und leiten die Testsitzungen. Das ist relativ teuer und von den Gruppendynamiken her nicht immer günstig. Wie sind bei Schüler/-innen unterschiedlichen Alters die Alternativen dazu einzuschätzen (besonders

die Durchführung durch Lehrpersonen an schuleigenen Computern)? Was sind die Gelingensbedingungen für Alternativen? Was für ein Testsetting führt zu hoher Testmotivation und Leistungsbereitschaft der Schüler/-innen, obschon keine individuelle Rückmeldung vorgesehen ist?

B3.18 Zu den ÜGK-Erhebungen wurden technische Berichte und Dokumentationen erstellt. Sind diese Berichte geeignet, die wissenschaftliche Güte der eingesetzten Instrumente aufzuzeigen?

B3.19 Für die ÜGK-Daten gibt es ein Datennutzungskonzept, das letztmals im Mai 2021 erneuert wird. Die in der ÜGK erhobenen Daten und die freigegebenen Instrumente werden (über FORSbase) der wissenschaftlichen Community zur Verfügung gestellt. Befriedigen das Datennutzungskonzept und die zur Verfügung gestellten Daten und Instrumente die Bedürfnisse der Bildungsforschung und der Testforschung?

Anpassung je nach Gebiet sinnvoll. Sie erinnert zudem daran, dass die Empfehlungen Nr. 1 bis Nr. 4 den Kindergarten / vorschulischen Bereich betreffen.

Zu Empfehlung Nr. 5

Die *Präsidentin* weist darauf hin, dass auch hier eine Unterscheidung nach Gebiet erforderlich ist.

Die *Generalsekretärin* entschuldigt sich für die kurze Konsultationsfrist, das Generalsekretariat ist fälschlicherweise davon ausgegangen, dass – wie bei der SODK – keine Konsultation bei den Mitgliedern notwendig sei. Die zeitliche Dringlichkeit ist auf die parallele Behandlung der parlamentarischen Initiative zurückzuführen. Sie ist dankbar, wenn die Frist eingehalten werden könne. Der politische Druck sei nämlich gross.

C. Schäli ergänzt, dass vor diesem Hintergrund eine kurze Frist sinnvoll ist.

Beschluss | *Décision*

Der Beschlussentwurf wird einstimmig gutgeheissen. | *Le projet de décision est accepté à l'unanimité.*

4. Überprüfung des Erreichens der Grundkompetenzen; Gesamtschau: Information und Aussprache | *Vérification de l'atteinte des compétences fondamentales; vision d'ensemble: information et discussion*

Die *Präsidentin* erinnert daran, dass die Plenarversammlung den Koordinationsstab HarmoS beauftragt hat, eine Gesamtschau zur Überprüfung der Grundkompetenzen zu erstellen. Aufgrund dieser Auslegeordnung sollen dann die weiteren Entscheide getroffen werden. Offenbar ist im Kosta HarmoS nun eine grundlegende Debatte entbrannt, weshalb die Generalsekretärin um eine Aussprache bittet. Die ÜGK ist seit Beginn ein umstrittenes Instrument. Einerseits soll damit die Entwicklung der Harmonisierungsbestrebungen der Kantone dokumentiert werden. Andererseits sollen die Erhebungen als Teil des Bildungsmonitorings die Qualitätsentwicklung vorantreiben. Nach den zwei ersten Durchgängen haben wir gesehen, dass mit dem heutigen Instrument eigentlich nur Kantonsvergleiche möglich sind. Das ist nicht befriedigend. Nun wird es darum gehen, aufgrund von Varianten zu entscheiden, wie es mit ÜGK weitergehen soll. Die Verwaltung muss wissen, in welche Richtung sie arbeiten soll. Das Ziel der heutigen Aussprache ist, dass wir der Generalsekretärin, die auch Präsidentin des Kosta HarmoS ist, eine Stossrichtung mitgeben können.

Die *Generalsekretärin* ruft mit Verweis auf Art. 62 Abs. 4 der BV sowie Art. 7 und 10 (Standards und Grundkompetenzen) des Konkordats die Grundlagen und Eckwerte der ÜGK in Erinnerung. Anlässlich der letzten Kosta HarmoS Sitzung hätte die Gesamtschau zuhanden des Vorstands verabschiedet werden sollen. Aufgrund der Diskussion im Kosta HarmoS stellte sich jedoch grundsätzlich die Frage, in welche Richtung die EDK gehen wolle.

Die *Präsidentin* fasst die zu klärenden Fragen zusammen: Wie ist die politische Haltung? Wie kommen wir auf vernünftige Daten, ohne das Schulsystem zu überhitzen? In welchem Umfang sind solche Erhebungen erforderlich?

C. Schäli betont, dass die Schweiz eines der teuersten Bildungssystemen weltweit hat. Gleichzeitig sind wir nicht im Stande die Frage zu beantworten, ob wird die Bildungsziele erfüllen. Es braucht daher eine solche nationale Bildungsmessung. Wenn wir «nein» zur ÜGK sagen, muss man sich auch bewusst sein, dass 20 Jahre Arbeit in den Sand gesetzt werden.

S. Kölliker betont, rückblickend müsse man nicht allzu viel ausbreiten. Er hat das Gefühl, dass zu Beginn Fehler gemacht wurden, insbesondere bei den Grundlagenarbeiten. Es wurde bisher mit viel Einsatz und Geld viel geleistet. Die Grundfrage der Leistungserhebungen stellt sich sowohl auf internationaler als auch auf nationaler Ebene. Im internationalen Vergleich empfindet er diese als nicht so wichtig. National haben sich Monitoringberichte etabliert. Diese werden sowohl von der Politik als auch von der Gesellschaft gefordert. Es braucht daher einen Vergleich. Die Auswahl der Varianten ist einzugrenzen. Ein massiver Ausbau kommt nicht in Frage. Dies könne man gegenüber der Bevölkerung nicht verantworten. Bisher war

die Aussagekraft klein. Mit Bezug auf ein vom Kanton Zürich zur Verfügung gestelltes Dokument kommt er zum Schluss, dass nur zwei Varianten denkbar sind: Eine minimale Anpassung der ÜGK oder den Vorschlag von *S. Wolter*, welcher eine Variante Pisa Plus vorschlägt.

C. Häsler unterstützt die vorgehenden Voten und dankt für die Gesamtschau, welche sehr hilfreich sei. Auch der Kanton Bern unterstützt den Vorschlag des Kantons Zürich. Die Erhebung ist für die Qualitätssicherung erforderlich. Es stellt sich jedoch die Frage nach dem Nutzen. Varianten mit massiven Kostensteigerungen kommen nicht in Frage. Aus Sicht des Kantons Bern kommen daher auch Variante 2 und 3 des Zürcher Papiers nicht in Frage. Sie betont, dass gute Anpassungen wichtig sind, das nächste Mal soll dies aber nicht im letzten Moment erfolgen. Den Kontextfragebogen hatte der Vorstand diskutiert, konnte dann aber keine Anpassungen beschliessen. Ob die Pisa-Variante eine Möglichkeit darstellt, muss mit Blick auf die Kosten geprüft werden. Die politische Diskussion über die Vermessung der Bildung entsteht immer wieder. Dem muss man Rechnung tragen. Daher müssen Alternativen sorgfältig abgewogen und geprüft werden. Der Kanton Bern kann sich mit Bezug auf das Zürcher Papier eine Variante 1 oder eine Variante Pisa Plus mit Stichproben, wie von *S. Wolter* vorgeschlagen, vorstellen.

C. Amarelle remercie pour le travail, l'audit et les recommandations faites. Elle rappelle que la vérification des compétences fondamentales (cofo) permet d'affiner le dispositif propre au canton. Ce monitorage fonctionne comme aiguillon d'amélioration de l'école au niveau romand et cantonal. Elle indique que grâce à la vérification des cofo, son canton a pu perfectionner son système d'enseignement en mathématiques. Elle se prononce en faveur de la variante 2 ou 3 avec les tests PISA qui resteraient complémentaires aux tests cofo. Elle tient à souligner que les cantons romands font figure de bons élèves en ce qui concerne les tests cofo et qu'en cas d'abandon ils se verraient dans l'obligation de refaire des épreuves communes romandes, ce qui seraient un véritable retour en arrière. Concernant les coûts, elle remarque que la Suisse peut se permettre une telle dépense et indique que ce n'est pas le système le plus cher si l'on compare à l'international.

M. Bertoli se dit toujours en faveur d'avoir des données comparables pour permettre au système de s'améliorer. Concernant les variantes, il souhaiterait des propositions plus détaillées. Il remarque que son canton a maintenu le système PISA afin d'avoir des comparaisons avec les pays proches. Que ce soit des comparaisons internes, ou avec l'utilisation de PISA ou PISA Plus, les deux propositions sont à son avis très bien.

M. Knill stellt fest, dass ihr das Dokument aus dem Kanton Zürich nicht vorliegt. Sie könne sich somit auf keine Varianten beziehen. Sie teile die Meinung, dass man auf die ÜGK nicht verzichten kann. In einer sich wandelnden Welt muss man sich überlegen, welchen Nutzen wir daraus ziehen. Der Kontext der Schule und der Schüler und Schülerinnen verändert sich auch aufgrund der Digitalisierung. Müsste man die Grundkompetenzen dynamischer gestalten? Anderenfalls müssten Unterschiede ständig erklärt werden. Es sei auch eine Überlegung wert, ob man nicht auf eine Onlineversion einschwenken könnte, analog zu Pisa, auch um Kosten zu sparen. Sie begrüsst daher irgendeine Form der Überprüfung, aber sieht noch nicht, was das Endprodukt sein sollte.

M. Courtet indique qu'il soutient les propos de ses collègues de la CIIP. Il se prononce en faveur de la variante 2 qui est ambitieuse, ou alors éventuellement la variante 3 mais dans tous les cas en maintenant les tests PISA.

Die *Präsidentin* stellt fest, dass alle der Meinung sind, es brauche ein Instrument zur Leistungsmessung. Sie habe kein Votum gehört, wonach die ÜGK beerdigt werden soll. Es seien einige kritische Stimme im Zusammenhang mit dem Kontextfragebogen entstanden und es besteht breiter Konsens darüber, dass die ÜGK optimiert werden muss. Es muss geklärt werden, ob beim Verfassen der Kontextfragebögen die politische Sicht gefragt ist. Sie ist der Meinung es handelt sich dabei um wissenschaftliche Fragen. Auch der Begriff der «Grundkompetenzen» muss neu definiert werden. Dies würde gemäss dem Zürcher Papier der Variante 1 entsprechen.

S. Kölliker fügt mit Verweis auf Variante 4 des Zürcher Papiers hinzu, dass diese Auslegung nicht dem HarmoS Konkordat entspricht. Er bitte darum, dass die Überlegungen von *S. Wolter* vertieft überprüft werden. Erst dann sei ein Beschluss möglich.

Die *Präsidentin* ergänzt, dass es sich bei Variante 1 des Zürcher Papiers um minimale Anpassungen der ÜGK handelt, Variante 2 sei ein Vorschlag der Universität Zürich und Variante 4 entspricht dem Vorschlag von *S. Wolter*, wobei hier noch das Problem der übrigen Stufen und der Fremdsprachen zu klären wäre.

F. Modoux betont als Mitglied der Kosta HarmoS, dass *S. Wolter* nicht die Variante 4 befürworte. Er habe vielmehr gesagt, dass Variante 2 die beste sei, aber auch die kostspieligste. Das Generalsekretariat hat aber noch keine genauen Berechnungen vorgenommen. Weiter habe er die Haltung vertreten, dass wenn man nicht konsequent in ÜGK investieren will, eine Erweiterung von Pisa in Frage käme. *S. Wolter* habe bedauert, dass die Schweiz so wenig investiert hat.

Die *Präsidentin* bedankt sich für diese Ergänzungen und bestätigt, dass sie *S. Wolter* auch so verstanden habe.

Die *Generalsekretärin* erinnert daran, dass es darum gehe, das weitere Vorgehen zu diskutieren. Die Varianten, die nun diskutiert werden, liegen ihr nicht vor und sie könne dazu nicht Stellung nehmen. Alle werden jedoch Kostenfolgen haben. Sie entnehme der Diskussion, dass diese Varianten substantiiert und den Mitgliedern vorgelegt werden sollen. Im Zusammenhang mit dem Kontextfragebogen werde sie einen Vorschlag unterbreiten.

Die *Präsidentin* betont, dass eine Mitgliedschaft im Kosta HarmoS die Verwaltung nicht von der Mitsprache entbinde. Eine solche Vernetzung ist erwünscht.

A. Gwerder weist darauf hin, dass heute eine Orientierung gelungen ist. Aus Sicht eines Kantons, der dem HarmoS Konkordat nicht beigetreten ist, steht er hinter der ÜGK. Der Kanton NW befürwortet zwei Reduktionen (Fächer und Stufen) und eine Erweiterung bei der Taktung (Testung).

Die *Präsidentin* weist darauf hin, dass im Generalsekretariat nur eine Person mit der Aufgabe betraut ist. Wenn detaillierte Analysen erforderlich sind, ist zu bedenken, dass das System überhitzt wird.

G. Küpfer bittet darum, bei der Überprüfung der Varianten den Nutzen in den Vordergrund zu stellen.

Die *Präsidentin* stimmt dem Votum zu. Die Steuerung sei ein wichtiger Punkt. Das Generalsekretariat muss sich von den Varianten lösen und andere erarbeiten können.

S. Kölliker fragt nach, wie nun der Auftrag für die weitere Zusammenarbeit laute. Es gab eine klare Mehrheit und nun ist der Varianten-Fächer wieder offen. Variante 2 ist offensichtlich keine Option.

Die *Generalsekretärin* betont, sie habe verstanden, dass eine Variante mit grossen Kostenfolgen keine Option sei. Die übrigen Varianten würden nun geprüft. Sie befürchtet, dass ÜGK ohne Mehrkosten gleichbedeutend ist mit dem Status quo. Sie weist auf die unterschiedlichen sprachregionalen Haltungen hin, die bei der Lösungsfindung berücksichtig werden müssen.

Auf die Rückmeldung von *S. Kölliker*, dass die Kantone minimale Anpassungen wollen, fügt die *Präsidentin* hinzu, dass die Anpassungen von *S. Wolter* nicht so minimal sind. Zum jetzigen Zeitpunkt ist man nicht entscheidungsfähig, das war auch nie das Ziel dieser Diskussion.

5. Teilnahme der Schweiz an PISA 2025: Beschlussfassung zuhanden der Plenarversammlung | *Participation de la Suisse à PISA 2025: décision à l'attention de l'Assemblée plénière*

Die *Generalsekretärin* führt in das Geschäft ein.

M. Knill unterstützt fast alle Anträge. Zur Option Englisch muss man sich jetzt im Klaren sein, ob man diese will oder nicht und was man daraus machen kann. Eine interessante Frage ist, ob unsere Kinder mit anderen Ländern mithalten können. Wenn wir unterdurchschnittlich sind, dann wird man Englisch ausbauen wollen zum Nachteil von Französisch. Das muss uns bewusst sein. Unsere Schulen sind durch Testungen schon genug belastet. Diese Option ist nicht hochprioritär, man kann darauf verzichten.

Die *Präsidentin* teilt diese Einschätzung. Nicht zu vergessen ist, dass es ein Schülersegment gibt für das Englischkenntnisse wichtig sind. Es ist auch die Sprache der Wissenschaft. Allerdings geht das zu Lasten unserer Mehrsprachigkeit.