

Checks im Kreuzfeuer der Kritik

Urs Moser stellt sich den Fragen des LVB



Schriftlich geführtes Interview mit Prof. Dr. Urs Moser vom Institut für Bildungsevaluation der Universität Zürich

Die standardisierten Leistungschecks riefen nach der ersten flächendeckenden Durchführung an den Primar- und Sekundarschulen im Bildungsraum Nordwestschweiz auf verschiedenen Ebenen Kritik hervor. In den Kantonen Baselland und Basel-Stadt sind bereits politische Vorstösse unterwegs, welche eine Reduktion der Checks fordern; die Konferenz der Lehrerinnen und Lehrer des Kantons Basel-Stadt forderte gar ihre Abschaffung. Wie interpretieren Sie diese breite Ablehnung?

Wir führen die Checks nun bereits seit vier Jahren durch. In jedem Kanton wurden wir zu Beginn mit vielen kritischen Fragen konfrontiert. Bald einmal erhielten wir auch positive Rückmeldungen. Die aktuelle Kritik ist wohl aufgrund der Tatsache entstanden, dass die Checks in den Kantonen Basel-Landschaft und Basel-Stadt zum ersten Mal auf der Sekundarstufe I durchgeführt wurden. Bevor an den Schulen der Mehrwert in Form von Ergebnisrückmeldungen sichtbar wird, fällt vor allem organisatorischer Mehraufwand an. Dieser Aufwand ist natürlich bei der ersten Durchführung besonders gross, weil alles neu ist. Bei der nächsten Durchführung wird der Aufwand deutlich geringer und der Mehrwert grösser sein, weil dann die Ergebnisrückmeldung für jeden einzelnen Schüler auch über die Zeit hinweg verglichen werden kann. Daher erwarte ich auch in den beiden Basel einen Stimmungswechsel.

Im Zuge der Aufgleisung des interkantonalen Bildungsmonitorings wurde immer wieder versichert, die standardisierten Checks dürften auf keinen Fall zu öffentlichen Rankings führen. Kritische Stimmen warnten dagegen seit jeher, es sei nur eine Frage der Zeit, bis es zu medial publizierten Rankings kommen

werde, was in anderen Ländern bereits vorgekommen ist. Wie gross ist Ihrer Meinung nach die Gefahr von Rankings und den damit einhergehenden Phänomenen «teaching to the test» sowie «naming, shaming and blaming»?

Ein Ranking soll und wird es aus mehreren Gründen nicht geben. Erstens wurden bereits vor der ersten Durchführung in allen vier Kantonen gesetzliche Massnahmen getroffen, die ein Ranking verbieten. Zweitens macht der Vergleich von Schulen keinen Sinn, weil die Checks jeweils nur in einer Klassenstufe durchgeführt werden und Schulen auf der Sekundarstufe I sehr unterschiedlich zusammengesetzt sind; manche umfassen alle Schultypen, manche zwei oder gar nur einen. Drittens sind die Checks ein Angebot für die Schulen, die alleine über ihre Ergebnisse verfügen, weshalb ein Ranking nur erstellt werden könnte, wenn die Schulen die Ergebnisse veröffentlichen würden. Dies aber ist weder im Sinn der Schulen noch erlaubt.

Der Lehrerschaft wurden die Checks ursprünglich als Teil eines Gesamtpakets schmackhaft gemacht, zu dem eine grosse, online verfügbare Aufgabensammlung gehörte. In dieser Hinsicht werden die Betroffenen seit Jahren vertröstet, während die Checks schon eingeführt wurden. Was können Sie hinsichtlich der Aufgabensammlung sagen?

Es war von Anfang an das Ziel der Kantone, die Aufgabensammlung so rasch wie möglich zur Verfügung zu stellen. Wir wollten aber nicht einfach Aufgaben ins Netz stellen, sondern die Aufgabensammlung für formatives, feedbackorientiertes Lernen nutzbar machen und auf den Lehrplan 21 abstimmen. Dabei war uns die wissenschaftliche Fundie-



FOTOLIA

Kein Grund zur Sorge

Urs Moser ist überzeugt, dass die Gefahr von Rankings und den damit einhergehenden Phänomenen «teaching to the test» sowie «naming, shaming, blaming» auch weiterhin nicht besteht.

rung besonders wichtig. Deswegen haben wir uns bei der Entwicklung vom renommierten Bildungsforscher John Hattie beraten lassen, der eine solche Aufgabensammlung für neuseeländische Schulen erfolgreich entwickelt hat. Die Umsetzung unserer Vision stellte sich dann als wesentlich aufwändiger heraus als erwartet.

Nun geht es endlich vorwärts. Anfang Mai 2017 haben wir die Schulen eingeladen, sich an der Erprobung der Aufgaben (Kalibrierung) zu beteiligen. Diese ist notwendig, um jedem Schüler und jeder Schülerin die ihren Voraussetzungen angemessenen Aufgaben anbieten zu können und um zuverlässige Rückmeldungen über den Lernfortschritt sowie über Stärken und Schwächen zu ermöglichen. Wenn sich die Schulen wie gewünscht an diesem Prozess beteiligen, sollte die Aufgabensammlung im Schuljahr 2017/18 zur Verfügung stehen.

Welchen Mehrwert – insbesondere unter dem propagierten Gesichtspunkt der Förderung – haben die Checks für die Lehrkräfte, wenn sie mit den ihnen zur Verfügung stehenden Daten nicht präzise rückverfolgen können, was genau ihre Schülerinnen und Schüler bei den einzelnen Aufgaben nicht schafften?

Die Förderung aufgrund von einzelnen Aufgaben ist nicht das Ziel der Checks. Ziel ist eine von der Lehrperson, der Klasse und vom Schultyp unabhängige Standortbestimmung. Die Ergebnisse können (1) förderorientiert im Vergleich zu Kompetenzstufen, (2) im sozialen Vergleich mit allen Schülerinnen und Schülern, (3) im individuellen Vergleich über die Zeit hinweg oder (4) auf der Sekundarstufe I im Vergleich zu den Anforderungsprofilen für die berufliche Grundbildung genutzt werden.

Bei einem Check ist die jeweils einzelne Aufgabe nicht so wichtig, denn es kann nie mit Sicherheit gesagt werden, aus welchen Gründen sie richtig oder falsch gelöst wurde. Die Checks prüfen deshalb anhand einer Vielzahl von verschiedenen Aufgaben und mit zunehmender Schwierigkeit verschiedene Kompetenzen, beispielsweise «komplexe Sachtexte, Erfahrungsberichte oder Zeitungsartikel verstehen und zentrale Inhalte aus dem Text erschliessen». Daraus entsteht ein objektives Gesamtbild.

Die Punktzahl in der Ergebnismeldung kann dann einer Kompetenzstufe zugeordnet werden, die beschreibt, was eine Schülerin oder ein Schüler beispielsweise im Lesen kann und welche Aufgabe sie oder er mit hinreichender Sicherheit richtig löst. Jede Kompetenzstufe wird unter Einbezug von Aufgabenbeispielen inhaltlich umschrieben. Der Mehrwert der Checks besteht also darin, dass sich die Ergebnismeldung auf mehrere Aufgaben bezieht, die die gleiche Kompetenz prüfen.

Der Check S2 gibt einerseits vor, formativ zu sein (daher die Anbindung an die leider noch nicht existierende Aufgabensammlung), andererseits werden die in ihm erzielten Ergebnisse für das Abschlusszertifikat der Volksschule verwendet, welches die bisherigen Checks der Wirtschaft (Basic Check und Multitick) ersetzen soll und einen klar summativen Wert hat. War Ihnen, als Ihr Institut den Check S2 ausgearbeitet hat, bekannt, dass eine solche Doppelnutzung angestrebt wurde, und wie schätzen Sie diese ein?

Ja, unbedingt, die Absicht war uns bekannt, und es gab deswegen auch Diskussionen. Der Check S2 kann förder-

orientiert genutzt und zugleich summativ interpretiert werden. Dies entspricht dem bekannten Spannungsfeld von fördern und beurteilen, mit dem die Schule seit jeher konfrontiert ist und das auch für die Checks von Bedeutung ist. Ich habe Verständnis für die Kritik der Doppelnutzung, weil die Checks primär auf die förderorientierte Nutzung und nicht auf die Selektion ausgerichtet sind.

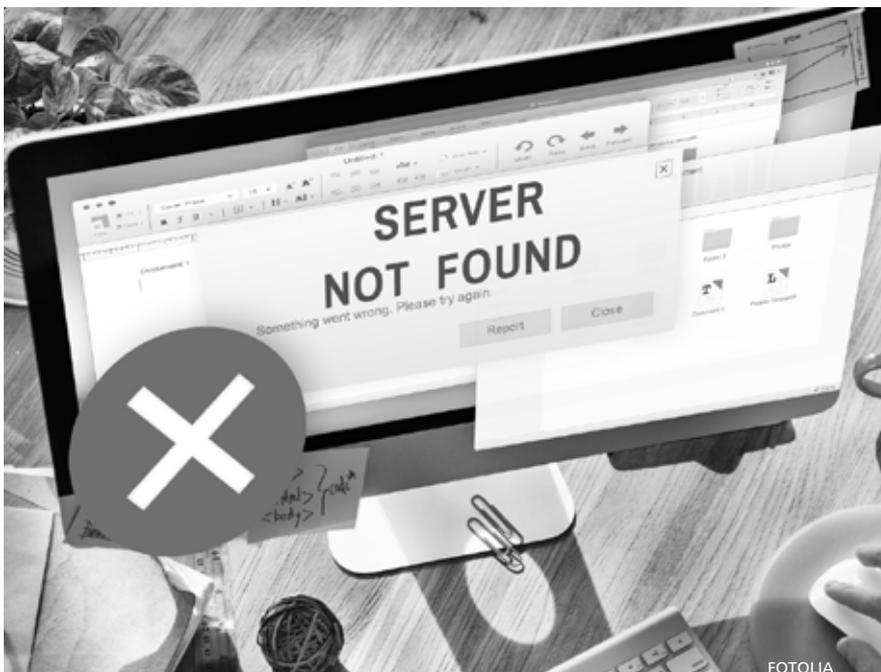
Zugleich bin ich aber dezidiert der Meinung, dass die Schule – insbesondere beim Übertritt in die berufliche Grundbildung – ihre Zuweisungsfunktion so gut wie möglich wahrnehmen muss, und dazu braucht es zwingend eine schultypenunabhängige Leistungsbeurteilung, wie sie der Check S2 ermöglicht. Für uns ist deshalb sehr wichtig, dass die Standardisierung der Durchführung und der Auswertung gegeben ist, sodass die Ergebnisse nicht willkürlich, sondern valide sind. Konkret heisst das: In allen Kantonen gilt das gleiche Zeitfenster für die Durchführung, alle Schülerinnen und Schüler lösen zu denselben Themen gleich viele Aufgaben und haben gleich lang Zeit dafür.

An vielen Schulen kam es zu grösseren technischen Problemen. Etliche Schülerinnen und Schüler mussten den Test wiederholen. Hinzu kamen mangelnde inhaltliche Passungen, so wurde beispielsweise in Chemie Wissen über Metalle abgefragt, das zuvor, gemäss Lehrplan, nie im Unterricht behandelt worden war. Halten Sie es für verantwortbar, bereits die erste flächendeckende Durchführung des Checks S2 für ein Abschlusszertifikat zu verwenden, das von den Betrieben, an denen sich die Schülerinnen und Schüler später bewerben, eingefordert werden kann?

Zu den technischen Problemen kam es, weil eine Besonderheit der Infrastruktur der Kantone Basel-Landschaft und Basel-Stadt nicht antizipiert wurde. In beiden Kantonen gelangen die Schulen über eine einzige IP-Adresse ins Netz, weshalb alle Schulen eines Kantons auf den gleichen Server geleitet, statt auf mehrere Server verteilt wurden. Wir konnten dieses Problem zum Glück im laufenden Betrieb beheben. Wir danken den Schulen, dass sie die dadurch irregulär durchgeführten Tests wiederholten. Bei der Auswertung haben wir darauf geachtet, dass irregulär durchgeführte Tests oder Testteile nicht in die Ergebnismeldung einfließen.

Den Vorwurf der mangelnden inhaltlichen Passung verstehe ich hingegen nicht. Wir haben mit der Schulpraxis gemeinsam ein Vorgehen konzipiert, wie der Check S2 Natur und Technik trotz fehlender curricularer Grundlage fair durchgeführt werden kann. Die Lehrpersonen konnten nach einem allgemeinen Test zu naturwissenschaftlichen Methoden aus sechs Tests zu Inhalten der Biologie, Chemie und Physik zwei auswählen. Es war also folglich möglich, den Check in Natur und Technik auch ohne Wissen über Metalle durchzuführen.

Uns wurde von der Schulpraxis versichert, dass mit diesem System für alle Schülerinnen und Schüler eine faire Situation gegeben sei und die Checks zu aussagekräftigen Ergebnissen führen werden. Das hat sich letztlich auch statistisch bestätigt. Wir werden den Check S2 Natur und Technik trotzdem im Juni 2017 mit der Schulpraxis und dem Zentrum Naturwissenschafts- und Technikdidaktik der PH FHNW evaluieren und optimieren.



Überlastung des Servers

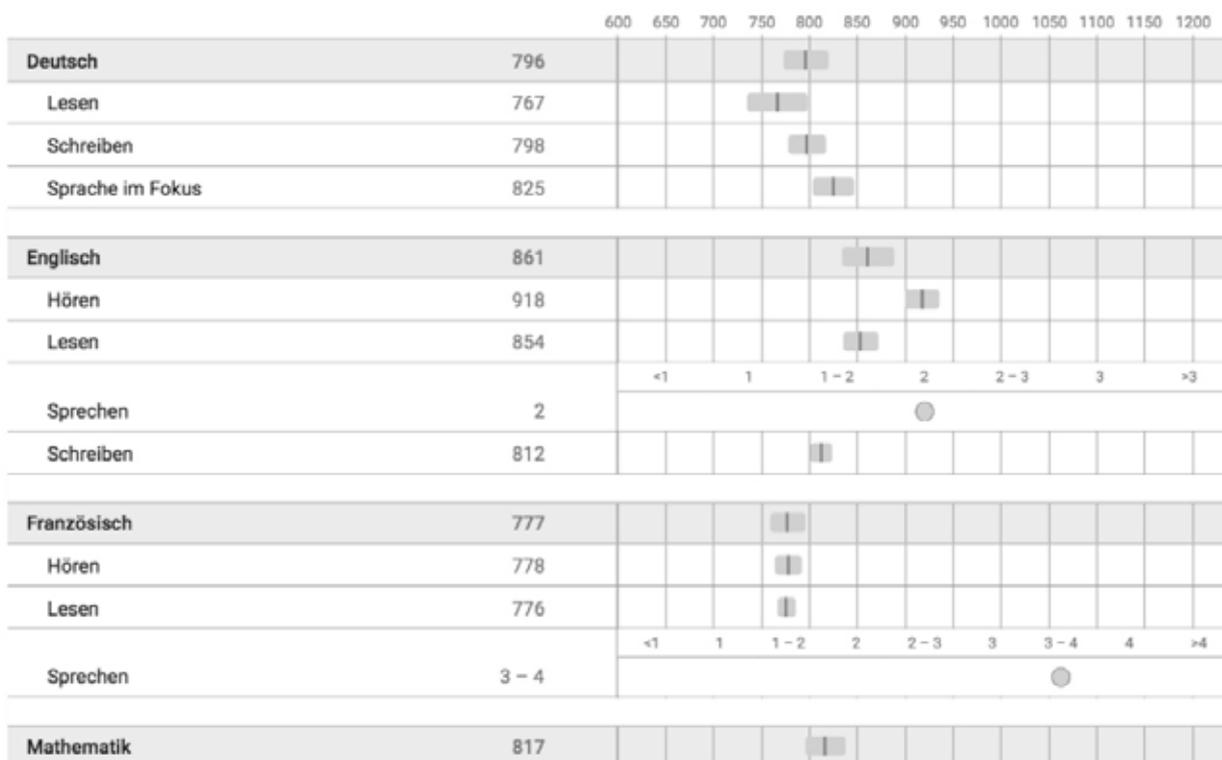
«Zu den technischen Problemen kam es, weil eine Besonderheit der Infrastruktur der Kantone Basel-Landschaft und Basel-Stadt nicht antizipiert wurde. In beiden Kantonen gelangen die Schulen über eine einzige IP-Adresse ins Netz, weshalb alle Schulen eines Kantons auf den gleichen Server geleitet, statt auf mehrere Server verteilt wurden.»

Die in den Checks erbrachten Leistungen der Schülerinnen und Schüler werden nicht in Form absoluter Zahlen ausgedrückt, sondern in relative Werte umgerechnet, die auf einer sogenannten Check-Skala zwischen 200 und 600 (Check P3), 400 und 800 (Check P6) respektive 600 und 1200 (Check S2) abgebildet werden. In den Kantonsberichten Ihres Instituts steht dieser Satz: «Ob eine Schülerin oder ein Schüler eine bestimmte Kompetenzstufe erreicht, sagt nichts darüber aus, ob die Schülerin oder der Schüler eine genügende oder ungenügende Leistung erbracht habe.» Gleichzeitig behaupteten Leserbriefschreiber aus der Bildungspolitik schlagartig, die Check-Ergebnisse seien ein Beleg für die (Un-)Wirksamkeit bestehender Schul- und Unterrichtsmodelle. Kritiker sehen sich durch dieses Wirrwarr an Aussagen darin bestätigt, dass diese Form der Leistungsabbildung einer «Blackbox» gleichkomme. Wir bitten Sie darum, anhand eines konkreten und auch für Laien verständlichen Beispiels unserer Leserschaft aufzuzeigen, wie diese Check-Skalen zustande kommen, und zu erläutern, welche Aussagen über das tatsächliche schulische Können der Schülerschaft sie ermöglichen.

Die Check-Skala beruht auf einem mathematischen Modell, das es erlaubt, die Aufgabenschwierigkeit und das Check-

Ergebnis auf der gleichen Skala abzubilden. Dazu wird die Beziehung zwischen Aufgabenschwierigkeit und Check-Ergebnis auf der Check-Skala berechnet. Mit anderen Worten ergibt sich die Wahrscheinlichkeit der richtigen Lösung einer Aufgabe als Funktion der Aufgabenschwierigkeit und der Fähigkeit eines Schülers. Das bedeutet: Je grösser die Fähigkeit ist, desto wahrscheinlicher ist es, dass ein Schüler eine bestimmte Aufgabe richtig löst. Und je schwieriger eine Aufgabe ist, desto unwahrscheinlicher ist es, dass ein Schüler mit einer bestimmten Fähigkeit die Aufgabe richtig löst. Es gibt meines Wissens kein besseres Konzept, um eine individuelle Leistung inhaltlich zu beurteilen, weil bei dieser Methode die Aufgabenschwierigkeit und die Fähigkeit gleichzeitig berücksichtigt werden.

Für die Checks werden nur Aufgaben eingesetzt, die diesem mathematischen Modell entsprechen, und dieses Modell überprüfen wir laufend mit den anfallenden Daten. Im Endausbau wird die Check-Skala von 200 bis 1200 Punkte reichen und sich vollständig durch Aufgaben illustrieren lassen. Jede Aufgabe lässt sich aufgrund ihrer Schwierigkeit auf der Check-Skala eindeutig verorten, gleich wie sich jedes Ergebnis eines Schülers auf der Check-Skala abbilden lässt. Um leichter interpretieren zu können, was die Punktzahl eines Schülers bedeutet, wird die Check-Skala in Kompetenzstufen unterteilt. Die Punktzahl kann einer Kompetenz-



«Die Check-Skala beruht auf einem mathematischen Modell, das es erlaubt, die Aufgabenschwierigkeit und das Check-Ergebnis auf der gleichen Skala abzubilden. Dazu wird die Beziehung zwischen Aufgabenschwierigkeit und Check-Ergebnis auf der Check-Skala berechnet.»

stufe zugewiesen werden, die darüber informiert, was der Schüler oder die Schülerin kann und welche Art von Aufgaben er oder sie mit hinreichender Sicherheit lösen kann.

Der Vorteil der Check-Skala ist, dass die Punktzahl immer positiv anhand von Kompetenzen und Aufgabenbeispielen umschrieben werden kann. Das Check-Ergebnis zeigt, was ein Schüler bereits sicher kann. Die Skala alleine sagt allerdings nichts darüber aus, ob die erbrachte Punktzahl genügt oder nicht genügt oder welche Note damit erreicht wird. Das inhaltsorientierte Vorgehen ist die Quintessenz der Kompetenzorientierung. Kompetenzorientiertes Lernen bedeutet erstens Lernen mit einem geschärften Blick auf Lernergebnisse und Lernfortschritte. Zweitens richtet es sich nach dem Aufbau des Wissens und Könnens, das in Kompetenzmodellen dargestellt ist. Und drittens orientiert es sich weniger an Jahrgangsziele und Noten als vielmehr am Lern- und Entwicklungsstand der Schülerinnen und Schüler. Ohne eine solche Skala, die übrigens auch der Aufgabensammlung zugrunde liegen wird, lässt sich ein kompetenzorientiertes Feedback nicht haben.

Eine Lehrperson wandte sich mit der folgenden Behauptung an uns: Wenn die leistungsstärksten Schülerinnen und Schüler in einem bestimmten Kompetenzbereich (z.B. Hörverständnis in Französisch) le-

diglich 60 von 100 möglichen absoluten Punkten erreichen, werden diese 60 Punkte in der Check-Skala dem Maximalwert zugeordnet. Dadurch entsteht ein irreführendes Bild über die tatsächlich erbrachten Leistungen in den Checks. Was entgegnet Sie?

Genau dieses irreführende Bild wird mit der Check-Skala verhindert: Das Ergebnis eines Schülers hängt in keiner Weise davon ab, wer am Check teilgenommen hat. Die Punktzahl auf der Check-Skala ergibt sich einzig aufgrund der Anzahl richtig gelöster Aufgaben und der Aufgabenschwierigkeit. Damit sind wir bei einem weiteren Vorteil der Check-Skala und dem mathematischen Modell, auf dem die Ergebnisse beruhen. Die Check-Skala bleibt über die Jahre hinweg konstant und misst immer gleich. Das bedeutet, dass die Punktzahlen aus zwei gleichen Checks, beispielsweise der Checks S2 in den Jahren 2016 und 2017, miteinander verglichen werden können.

Auch die Ergebnisse des Checks S2 und des Checks S3 werden auf der gleichen Skala abgebildet. Durch den Vergleich der beiden Ergebnisse wird der individuelle Lernfortschritt zwischen den beiden Checks ersichtlich (individuelle Bezugsnorm). Aus diesem Grund ist die Check-Skala auch gegen oben und unten nicht beschränkt. Sie könnte theoretisch auch Werte geringer als 200 oder Werte grösser als



Abgestufte Hörverständnis-Kompetenzen in Französisch und Englisch für angehende Binnenschiffahrtsmatrosen: Sind Anforderungsprofile ein Produkt der Vermessungseuphorie?

FOTOLIA

Laut Moser stützt sich das Konzept auf «anerkannte psychologische Theorien». «Allerdings beruhen die Anforderungsprofile bis anhin ausschliesslich auf Experteneinschätzungen, was noch nicht garantiert, dass sie in jedem Fall valide sind.»

1200 annehmen. Die Check-Skala mag etwas ungewohnt sein, sie hat aber den Vorteil, dass die Punktzahl eine inhaltliche Bedeutung hat und deshalb zu einem aussagekräftigen Bild über die tatsächlichen Leistungen führt.

Nach Auswertung des Checks S2 kann sich jede Schülerin und jeder Schüler auf www.check-deinwissen.ch einloggen und herausfinden, ob ihre respektive seine Checkresultate mit den Anforderungen der jeweiligen Wunschberufe in Einklang stehen. Gemäss Website kommen die Anforderungsprofile aufgrund von Experteneinschätzungen zustande. Macht man die Probe aufs Exempel und klickt sich einmal durch ganz verschiedene Anforderungsprofile, gewinnt man allerdings den Eindruck, die Verantwortlichen hätten vor lauter Vermessungseuphorie den nüchternen Blick auf die Anforderungen der Berufswelt verloren (z.B. abgestufte Hörverständnis-Kompetenzen in Französisch und Englisch für angehende Binnenschiffahrtsmatrosen oder Bühnentänzerinnen?). Für wie alltagstauglich halten Sie diese Auflistungen?

Das Konzept beruht auf anerkannten psychologischen Theorien, welche die Passung zwischen Fähigkeiten und Anforderungen als Voraussetzung für eine erfolgreiche berufliche Entwicklung ansehen. Wir haben zudem den Abgleich zwischen den Check-Ergebnissen und den Anforderungsprofilen Anfang Jahr systematisch analysiert und gewisse notwendige Anpassungen bei der Transformation der Check-Ergebnisse auf die Anforderungsprofile vorgenommen. Der Vergleich der Check-Ergebnisse mit den Anforderungsprofilen verschiedener beruflicher Grundbildungen kann sowohl für die Orientierung und Reflexion im Laufe des Berufswahlprozesses als auch für die Planung des Lernens als praxistauglich bezeichnet werden. Es ist unseres Erachtens gleichwohl von Vorteil, wenn die Ergebnisse mit zusätzlichen Informationen und gemeinsam mit einer Fachperson interpretiert werden. Insgesamt erachten wir das Instrument als eine gute Grundlage, in der Entwicklung aber noch längst nicht abgeschlossen.

Dass die Einschätzung der Experten nicht immer mit der persönlichen Einschätzung übereinstimmt, heisst jedoch nicht, dass die Experten falsch liegen. Allerdings beruhen die Anforderungsprofile bis anhin ausschliesslich auf Experteneinschätzungen, was noch nicht garantiert, dass sie in jedem Fall valide sind. Deshalb ist es unseres Erachtens wichtig, die Check-Ergebnisse mit dem Erfolg in der beruflichen Grundbildung in Verbindung zu bringen und die Validität der schulischen Anforderungsprofile zu überprüfen. Zudem müssten die Anforderungsprofile unbedingt an den Lehrplan 21 angepasst werden.

Betreibt auch Ihr Institut eine Art Manöverkritik? Können Sie bereits Aussagen dahingehend treffen, welche Aspekte der Checks aus Ihrer Sicht optimiert werden müssten?

Das Bessere ist der Feind des Guten, sagt Voltaire, und ich gebe ihm Recht. Ein Vorhaben dieser Grössenordnung entwickelt sich erfahrungsgemäss durch Kritik, die aus unterschiedlicher Perspektive erfolgt, auch aus einer wissenschaftlichen Perspektive und vor allem durch Selbstkritik unsererseits. Aufgabensammlung, Checks in Natur und Technik, Check-Ergebnisse und Anforderungsprofile befinden sich in einem Entwicklungsstadium, das zwar fortgeschritten, aber noch nicht abgeschlossen ist. Unsere Dokumentation der wissenschaftlichen Methoden und der Testgütekriterien sind noch im Aufbau. Auch unsere Kommunikation haben wir noch zu optimieren, damit Funktion und Nutzen der Checks, aber auch der Umgang mit den Ergebnissen für alle Beteiligten klar und verständlich sind.

Wie lautet Ihre Botschaft an Politik, Lehrerschaft und Öffentlichkeit, was die Checks betrifft?

Gehen wir einmal davon aus, dass die Checks und Aufgabensammlung so wie geplant umgesetzt werden, also mit curricular validen Aufgaben und ohne unerwünschte Nebenwirkungen wie «teaching to the test», dann sind die Ergebnisrückmeldungen ohne Zweifel eine wichtige Ergänzung der eher subjektiven Leistungsbeurteilung durch Noten im Schulalltag. In diesem Sinne liefern sie einen pädagogisch wertvollen Beitrag zu einer fairen Beurteilung und zu einer datengestützten Planung des Lehr-Lern-Prozesses. Es geht eben nicht darum, einzelne Lehrpersonen zu kritisieren, sondern darum, sie zu unterstützen. Die Nutzung eines unabhängigen Feedbacks über den Lernstand der eigenen Klasse gehört zum professionellen Handeln einer Lehrperson, die auf Stärken und Schwächen ihrer Schülerinnen und Schüler eingeht.

Die Auseinandersetzung mit dem Leistungsstand der Schülerinnen und Schüler gehört aber auch zum Qualitätsmanagement von Schulen. Die Checks liefern relevante und wissenschaftlich abgesicherte Informationen für die Schulentwicklung und ermöglichen Transparenz über den Umgang mit sozialen Ungleichheiten. Mit den Checks und der Aufgabensammlung geht es in keiner Weise um die Vermessung von Bildung, sondern um die Ergänzung der schul- und klassenbezogenen Beurteilung ausgewählter Leistungen im Interesse von Schülerinnen und Schülern, Lehrpersonen und Schulen.