

Frau  
Präsidentin des Nationalrates  
Mag. Barbara Prammer  
Parlament  
1017 Wien

XXIV. GP.-NR

11634 /AB

07. Aug. 2012

Geschäftszahl:

BMUKK-10.000/0182-III/4a/2012

zu 11803 /J

Wien, 2. August 2012

Die schriftliche parlamentarische Anfrage Nr. 11803/J-NR/2012 betreffend Durchführung der Bildungsstandards-Testung in Mathematik am 23. Mai 2012, die die Abg. Hermann Gahr, Kolleginnen und Kollegen am 6. Juni 2012 an mich richteten, wird wie folgt beantwortet:

Zu Frage 1:

Die Bildungsstandards stellen ein wichtiges Instrument der Qualitätssicherung im Bildungsbereich dar. Die Hauptzielrichtung ist dabei, einen Beitrag zu einer systematischen, insgesamt für Schülerinnen und Schüler förderlichen Schul- und Unterrichtsentwicklung zu leisten. Unter Hinweis auf § 17 Abs. 1a des Schulunterrichtsgesetzes „... ist [es] vorzusehen, dass die Ergebnisse von Standardüberprüfungen so auszuwerten und rückzumelden sind, dass sie für die langfristige systematische Qualitätsentwicklung in den Schulen nutzbringend verwertet werden können.“

Die Überprüfung erfolgt flächendeckend, damit alle Schulen eine Rückmeldung darüber bekommen, wie weit es ihnen gelungen ist, ihre Schülerinnen und Schüler zum angestrebten Kompetenzstand zu führen. Die Schulaufsicht, die Schulbehörden und das Bundesministerium für Unterricht, Kunst und Kultur erhalten wesentliche Steuerungsinformationen.

Zu Fragen 2 und 3:

Es wurden keine Feldtestungen in Mathematik vorgenommen, es gibt aber Pilotierungen zur Prüfung der Itemqualität und 2009/10 gab es Baseline-Untersuchungen (Ausgangsmessungen) in großen Stichproben von Schülerinnen und Schülern der 4. und 8. Stufe. Diese Schülerinnen und Schüler wurden noch nicht nach den Standards unterrichtet – deswegen auch „Ausgangsmessung“. Die Ergebnisse werden ab 2012 als Bezug für Längsschnittsdarstellungen dienen.

Vergleichende Auswertungen wurden nicht beauftragt. Die Ergebnisrückmeldungen an Schülerinnen und Schüler, Lehrerinnen und Lehrer sowie Schulleitungen dienten primär zur Erprobung der Verfahren.

Zu Frage 4:

Anliegen ist es, Vergleiche in einer Weise anzustellen, die konstruktive Möglichkeiten der Qualitätsverbesserung eröffnen. Die Kriterien Schulart, städtischer/ländlicher Bereich (ua.) fließen in die Extrainformation „fairer Vergleich“ ein.

Die Hauptzielrichtung der Bildungsstandardüberprüfung ist Qualitätsentwicklung: Jede Lehrkraft, jede Schule, die Verantwortlichen im System sollen sich mit ihren Ergebnissen konstruktiv auseinandersetzen und ihre Qualität steigern – wenn das Ergebnis für sie erfreulich war, sollen sie die förderlichen Faktoren absichern und allfällige Schwachstellen in Angriff nehmen; wenn das Ergebnis unerfreulich war, sollen sie die Ursachen dafür identifizieren und einen Plan zur schrittweisen Verbesserung entwickeln.

Der Vergleich zwischen Schulararten, Regionen und Ähnliches bringt hauptsächlich Effekte von Rahmenbedingungen in den Blick, die von den Betroffenen nicht oder nur bedingt beeinflussbar sind. Das ist weder im einen noch im anderen Fall förderlich dafür, die eigene Verantwortung wahrzunehmen und – ausgehend vom eigenen jeweils aktuellen Stand – die Qualitätsentwicklung weiter voranzubringen.

#### Zu Frage 5:

An der Überprüfung haben nur die Klassen der 8. Schulstufe teilgenommen. Zu diesem Zeitpunkt gab es vorerst 67 Neue Mittelschule(NMS)-Standorte mit 171 Klassen auf der 8. Schulstufe. Dazu kommt, dass es sich bei diesen NMS-Standorten um jene handelt, die als erste (Generation 1) in den Modellversuch eingestiegen sind – die Ergebnisse dieser NMS-Standorte sind aufgrund geringer bzw. nicht vorhandener Fallzahlen (nur 5 Bundesländer in Generation 1) nicht repräsentativ.

#### Zu Frage 6:

Für die ersten beiden Generationen der NMS werden die Daten der Bildungsstandardsüberprüfungen herangezogen, um einerseits längsschnittliche Analysen anzustellen und andererseits die NMS-Absolventinnen und -Absolventen der Vergleichsgruppe gegenüberzustellen. Da die Überprüfung der Bildungsstandards 2012 nur Mathematik und 2013 nur Englisch umfassen, wurden sie in erweiterten Testsitzungen für die NMS ergänzt durch Kurzskalen für Deutsch und Englisch (Generation 1, 2012) bzw. Mathematik (Generation 2, 2013).

#### Zu Frage 7:

Bei der Annahme, die Testergebnisse der Schülerinnen und Schüler würden für die Ergebnisberichterstattung verändert, handelt es sich um ein grundlegendes Missverständnis. Das rückgemeldete Testergebnis jeder Schülerin/jedes Schülers Schülerin hängt ausschließlich davon ab, wie viele der 48 Aufgaben im jeweiligen Testheft gelöst wurden. Dieses Ergebnis wird rückgemeldet und bildet die Grundlage für den Vergleich der absoluten Ergebnisse.

Der „faire Vergleich“ bietet eine zusätzliche Information, um die Aussagekraft des Ergebnisses zu erhöhen: Das Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens (BIFIE) berechnet einen „Erwartungswert“, der die unterschiedlichen Rahmen- und Kontextbedingungen der Schülerinnen und Schüler sowie der Schulen berücksichtigt (zB. Bildungshöhe aller Eltern am Standort, Anteil von Kindern nichtdeutscher Muttersprache, Urbanisierungsgrad, Schulgröße). Der Erwartungsbereich einer Schule ist dann jener Leistungsbereich, der aufgrund der gegebenen strukturellen Rahmenbedingungen zu erwarten wäre. Für Schulen mit vergleichbaren strukturellen Rahmenbedingungen würde sich der gleiche Erwartungsbereich ergeben.

Die Ergebnisse werden jedenfalls unverändert berichtet, und sie werden in zweierlei Hinsicht verglichen: mit einem absoluten Vergleichswert (zB. Österreich-Schnitt) und zusätzlich mit dem

„fairen“ Erwartungswert (Ein Muster für den Ergebnisbericht ist im Internet bereits abrufbar: <https://www.bifie.at/node/64> ). In die Evaluation gehen alle diese Informationen ein.

Zu Frage 8:

Für die NMS-Evaluation wurden eigene Ausgangsmessungen durchgeführt. So wurden zur Vergleichsgruppenanalyse im Schuljahr 2008/09 bzw. 2009/10 die letzten Schülerinnen und Schüler der 8 Schulstufe, die die neuen NMS-Standorte noch im damaligen Regelsystem besucht hatten, mit Kurzskalen der Bildungsstandards getestet. Die damals erhobenen Daten werden mit den jetzigen Daten der Bildungsstandardsüberprüfung verglichen, wobei die NMS-Standorte für sich selbst die Vergleichsgruppe darstellen, um die Frage zu beantworten, was sich auf Schülerebene seit der Einführung des Modells NMS geändert hat.

Zu Frage 9:

Es erfolgt keine Veränderung der Ergebnisse der Schülerleistungen; Auf die Beantwortung der Frage 7 wird verwiesen. Im multiplen Regressionsmodell zur Ermittlung der Erwartungswerte werden derzeit auf Schulebene folgende Kontextvariablen einbezogen: Schulgröße/Gruppengröße, Gemeindegröße, Schulart (HS, AHS, NMS), Schulerhalter (öffentliche, privat), Urbanisierungsgrad und Entfernung zur nächstgelegenen AHS, Anteil der Mädchen/Burschen, Anteil der Schülerinnen und Schüler mit/ohne Migrationshintergrund und Anteil der Schülerinnen und Schüler, deren Erstsprache Deutsch ist, Sozialstatus der Schülerinnen und Schüler (Ausbildung und beruflicher Status der Eltern, Anzahl der Bücher zu Hause), Anzahl der von der Testung ausgenommenen Schülerinnen und Schüler.

Zu Frage 10:

Die Gewichte (sog. Beta-Werte oder Regressionskoeffizienten) entstehen aufgrund eines multiplen Regressionsmodells, das aus den Daten der am 23. Mai 2012 getesteten Schülerinnen und Schüler ermittelt werden wird.

Zu Frage 11:

Das Ergebnis einer Schülerin/eines Schülers wird für den fairen Vergleich nicht korrigiert. Die Auswirkung einer bestimmten Konstellation von Kontextfaktoren für den Erwartungswert kann erst berechnet werden, wenn das laufende Datenmanagement für die mehr als 80.000 Schülerinnen und Schüler aus der aktuellen Überprüfung und die Analysen dieser Datensätze durchgeführt worden sind. Ein genereller Wert für „AHS/Kleinstadt“ wird allerdings nicht berechnet, sondern ein volles Modell für jeden Schulstandort.

Zu Frage 12:

Analog zur Beantwortung der Frage 11.

Zu Frage 13:

Die Ergebnisse werden nicht hochgerechnet. Sie werden lediglich zu zwei Vergleichsbasen in Beziehung gesetzt.

Der faire Vergleich soll – zusätzlich zum Vergleich der absoluten Ergebnisse – zeigen, welches Ergebnis in den gegebenen Rahmenbedingungen erreichbar ist. Damit soll sich der Blick öffnen auf die tatsächliche Leistung der Schule/Lehrkraft, auf jene Qualität, die tatsächlich selbst geschaffen ist und nicht der Gunst oder Ungunst äußerer Rahmenbedingungen zu danken ist.

Analoge Vorgangsweisen sind in fast allen europäischen Staaten bei Standardüberprüfungen üblich.

Zu Frage 14:

Die standortbezogene Qualitätsentwicklung ist stark fokussiert auf innerschulische Prozesse und Kompetenzen. Fragen der Schulqualität müssen umfassend auf allen Ebenen, auch unter Berücksichtigung des Nationalen Bildungsberichts, internationaler Assessments und diverser nationaler Evaluierungen geführt werden.

Zu Frage 15:

Der Bundesergebnisbericht wird zum einen absolute Schülerergebnisse nach erreichten Leistungen und zum anderen nach erreichten Kompetenzstufen (kriteriale Rückmeldung) enthalten. Auf Bundesebene rückt der faire Erwartungswert in den Hintergrund.

Zu Frage 16:

Es gibt unterschiedliche Testheftformen, sie sind jedoch für sich genommen gleich schwierig. Es kann allerdings sein, dass einer Schülerin/einem Schüler einzelne Aufgaben eher „liegen“ als andere. Dem soll das Vertrauensintervall Rechnung tragen.

Jede Testung – egal wie detailliert – kann immer nur einen Ausschnitt der tatsächlichen Leistungsfähigkeit abbilden. Die aktuelle Überprüfung ist so konzipiert, dass in einem für die einzelnen Schülerinnen und Schüler zumutbaren Zeitausmaß möglichst viele inhaltliche Aspekte abgedeckt werden. Die Treffergenauigkeit steigt mit der Anzahl der Schülerinnen und Schüler: 48 Items je Testheft lassen für die/den einzelne/einzelnen Schülerin/Schüler nur eine relativ genaue Rückmeldung des Ergebnisses zu. Auf Klassenebene ergänzen die Testhefte einander, sodass die Genauigkeit der Rückmeldung steigt. Auf Schulebene steigt sie nochmals, usw. Um diesen Effekt zu erzielen, müssen unterschiedliche Testheftformen eingesetzt werden:

Die hier zitierten Werte sind fiktiv. Sie stammen aus einem Musterheft, das veranschaulichen soll, welche Aspekte die Rückmeldungen erfassen werden.

Zu Frage 17:

Das Vertrauensintervall hängt ua. von der Gruppengröße und der Homo- bzw. Heterogenität der Gruppe ab. Bei der Baselinetestung betrug das Vertrauensintervall bei einer Gruppe von 22 Schülerinnen und Schüler (Klassenstärke) rund 22 Punkte, bei einer Gruppengröße von 50 Schülerinnen und Schüler (mittelgroße Schule) rund 14 Punkte.

Zu Frage 18:

Nach dem Zufallsprinzip: Es gibt mehrere zufällig rotierte Testhefte. Es wurde beim Testen nicht nach den Schultypen unterschieden, die Tests waren für alle Schulen und Schultypen vom gleichen Schwierigkeitsgrad.

Zu Frage 19:

Die Bildungsstandards „basieren auf grundlegenden Kompetenzen, über die die Schülerinnen und Schüler bis zum Ende der jeweiligen Schulstufe in der Regel verfügen sollen“ (§ 2 Z 1 der Verordnung über Bildungsstandards im Schulwesen). Dies impliziert, dass eine große Anzahl von Schülerinnen und Schülern die getesteten Inhalte so gut beherrscht, dass sie die Testaufgaben als „sehr leicht“ empfanden. Wie bei jedem Test wird es allerdings auch eine

gewisse Anzahl an Schülerinnen und Schüler geben, denen die Lösung der Aufgaben schwer gefallen ist.

Die Intention der Bildungsstandards ist nicht, eine Hürde zu bilden, die nur eine beschränkte Anzahl von Personen bewältigen kann. Die Intention ist, dem Lehren und Lernen eine objektive, gesellschaftlich akkordierte Orientierung zu geben: „Der Lehrer hat bei der Planung und Gestaltung seiner Unterrichtsarbeit die Kompetenzen und die darauf bezogenen Bildungsstandards zu berücksichtigen sowie die Leistungen der Schüler in diesen Bereichen zu beobachten, zu fördern und bestmöglich zu sichern.“ (§ 17 Abs. 1a des Schulunterrichtsgesetzes).

#### Zu Frage 20:

Die Kosten für den ersten Überprüfungszyklus 2012 bis 2014 betragen EUR 23,5 Mio. Dies umfasst sowohl Testkosten, Kosten für die Rückmeldungen, als auch Kosten für die begleitenden Implementierungsmaßnahmen in der Lehrerfortbildung. Davon sind im Jahr 2012 für Überprüfung und Implementierung EUR 5,8 Mio. budgetiert.

2012 kostete der Test/Schülerin bzw. Schüler EUR 68,6.

Der erste Überprüfungszyklus umfasst insgesamt 5 flächendeckende Testungen auf der 4. und 8. Schulstufe, an denen ca. 430.000 Schülerinnen und Schüler teilnehmen werden.

Unter Implementierungskosten sind ua. zu subsumieren:

- die Information aller beteiligten Personengruppen und Systemebenen, soweit vom BIFIE durchgeführt,
- die Publikation und fortlaufende Aktualisierung unterrichtsnaher Lehr-, Lern- und Begleitmaterialien (Praxishandbücher, Themenhefte, Aufgabensammlungen, Leitfäden, Broschüren, etc.) in gedruckter bzw. elektronischer Form,
- Informations- und Fortbildungsveranstaltungen für Multiplikatorinnen und Multiplikatoren, Führungskräfte (LSI, BSI, Schulleitungen) und Lehrende an Pädagogischen Hochschulen, soweit vom BIFIE durchgeführt,
- die Entwicklung und Betreuung von Diagnoseinstrumenten zur Informellen Kompetenzmessung (IKM), die Lehrkräften Lernstandserhebungen auf Schüler- und Klassenebene ermöglichen und damit wesentlich zur gezielten Förderung der Lernenden im Unterricht beitragen.

Die Bundesministerin:

