

Standardisierung, Testitis und die Amputation von Bildung

Oder: Ein Plädoyer gegen die Normierung und Vermessung des Menschen. Eine Streitschrift.

Jörg Spenger¹

Zusammenfassung

Internationale Vergleichsstudien haben ein gemeinsames und vorrangiges Ziel, nämlich die Kompetenzen bzw. Leistungen von Schülerinnen und Schülern in ausgewählten Domänen und auf unterschiedlichen Schulstufen aus den Teilnehmerländern miteinander zu vergleichen, um etwa den einzelnen Bildungssystemen Rückmeldungen über die jeweilige Performance zu geben und potentielle Verbesserungen aufzuzeigen bzw. anzuregen.

Gleiches gilt an sich für die nationalen Bildungsstandards, die im Sinne eines System-Monitorings die Leistungen der Schülerinnen und Schüler innerhalb Österreichs überprüfen und z.T. auch den Anspruch erheben, für mehr Bildungsgerechtigkeit und Vergleichbarkeit zu sorgen, zumindest was die teilzentrale Reife- und Diplomprüfung betrifft.

Diese Testungen haben z.B. eine Reihe von Defiziten des österreichischen Schulsystems zu Tage gefördert, die bereits Ansatzpunkt für notwendige bildungspolitische Maßnahmen waren, sind und noch weiter sein werden. Die Grenzen dieser Testungen liegen – neben testtheoretischen Limitationen – besonders dort, wo sie den Anspruch erheben, Bildung oder den „Erfolg“ eines Schulsystems (was auch immer das ist) insgesamt zu messen, denn diesen lösen sie nicht ein, da nur ein sehr schmaler Ausschnitt dessen getestet wird, was allgemein als Bildung bezeichnet wird bzw. zumindest dessen, was an unseren Schulen in den einzelnen Fächern unterrichtet wird und auch werden soll. Die große Gefahr besteht deshalb u.a. in einer (bewussten oder unbewussten) Fehlinterpretation der Daten und in entsprechenden (unangemessenen oder überzogenen) bildungspolitischen Entscheidungen.

Ein großes Problem ist – neben anderen – auch das bekannte Phänomen „Teaching to the test“; dass also besonders das im Unterricht favorisiert wird, was getestet wird, und dass jene Gegenstände oder Lernbereiche mehr oder weniger „ausgeblendet“ oder zumindest vernachlässigt werden, die nicht testrelevant sind. Es kann also dazu kommen, dass der Unterricht an den Schulen besonders jene Kompetenzen in den Blick nimmt, die überprüft werden, was in weiterer Folge einerseits zu einem ziemlich „schmalspurigen“ Verständnis von Bildung führen könnte, und andererseits dazu, dass viele wertvolle Domänen des Lehrplans einfach „links liegen gelassen“ werden. Bildung in ihrer ganzen Breite standardisieren oder gar messen zu wollen, scheint ohnehin ein unerfüllbares, illusionäres Wunschdenken zu sein.

Die vorliegende Arbeit versucht, nationalen und internationalen Schulleistungstests auf die Spur zu kommen und ihre Möglichkeiten und Grenzen auszuloten. Dabei wird zunächst deren historische Entwicklung kurz beleuchtet, um in einem weiteren Schritt die Vor- und Nachteile, aber auch mögliche Gefahren, die von Testungen wie PISA, TIMSS, PIRLS, Bildungsstandards, „Zentralmatura“ & Co ausgehen, zu diskutieren. Außerdem werden mögliche und notwendige Konsequenzen aus den Ergebnissen der Vergleichsstudien abgeleitet, insbesondere für Österreich. Abschließend wird jedoch auch auf die Gefahren der Normierung von Unterricht durch Tests und Leistungsvergleiche hingewiesen und generell dem Versuch, Bildung standardisieren und messen zu wollen, eine Absage erteilt.

Schlüsselwörter:

Standardisierung
Testung
Bildung

¹ Pädagogische Hochschule Niederösterreich, Mühlgasse 67, 2500 Baden.
Korrespondierender Autor. E-Mail: joerg.spenger@ph-noe.ac.at

1 Einleitung

Das Thema „Internationale und nationale Schulleistungsvergleiche bzw. ihre Möglichkeiten und Grenzen“ interessiert mich schon seit Langem; insbesondere als jemand, der auch selbst noch einige Stunden an einer Schule unterrichtet, d.h. in der Praxis steht, lässt es mich nicht kalt, da in den letzten Jahren geradezu ein „Evaluations- und Überprüfungswahn“, sei es national oder international begründet, eingesetzt und Schule und Unterricht – man möchte fast sagen – dramatisch verändert hat. Besonders problematisch ist allerdings aus meiner Sicht der sehr unreflektierte Umgang der (Schul-)Politik, einiger Medien und leider auch nicht weniger Vorgesetzter im Schulsystem mit dieser Thematik. Aus meiner Sicht hat das im letzten Jahrzehnt beinahe mehr Schaden angerichtet als Nutzen gebracht und der Sache keinen guten Dienst erwiesen.

An dieser Stelle möchte ich deshalb auch die Gelegenheit ergreifen, mich bei jenen Personen zu bedanken, die mich zum Schreiben der vorliegenden Arbeit erst motivierten: Danke, liebe Politikerinnen und Politiker, Journalistinnen und Journalisten, Schulaufsichtsbeamtinnen und -beamte, Direktorinnen und Direktoren, dass ihr mich mit euren Worten und Taten angeregt habt, diese Zeilen zu verfassen. Dabei sei natürlich auch und insbesondere jenen aus den oben erwähnten Personengruppen gedankt, die sehr verantwortungsvoll mit der Thematik umgehen. Gott sei Dank gibt es Sie auch!

Es wird aus dem Gesagten sehr schnell klar, dass dieser Text natürlich subjektive Betrachtungen der Situation liefert, jedoch damit bewusst wachrütteln möchte. Er versucht allerdings insofern ausgewogen zu sein, als das Thema aus verschiedenen Blickwinkeln betrachtet wird und sich stets auf wissenschaftlicher Basis bewegt. Dies ist auch meiner Arbeit als Bildungsforscher an Universitäten und Hochschulen geschuldet. Im Rahmen dieser Tätigkeit beschäftige ich mich seit Jahren in Lehre und Forschung mit der Thematik und stelle immer wieder fest, wie oberflächlich manche Argumentationen häufig sind. Dem soll hiermit entgegengetreten werden.

1.1 Erkenntnis- und Praxisinteresse, pädagogische Relevanz

Die anfängliche Serie von Pannen bei der Abwicklung der so genannten „Zentralmatura“ beim Bundesinstitut für Bildungsforschung (BIFIE) (z.B. Nimmervoll, 2014) hat den allgemeinen Diskurs neu befeuert, welche Möglichkeiten und Grenzen zentrale, nationale und internationale Testungen, wie PISA, PIRLS, TIMSS, Bildungsstandards oder die neue standardisierte kompetenzorientierte Reifeprüfung, haben. Nicht nur von Lehrerseite und insbesondere von Seiten der Lehrgewerkschaft (Schmid, 2014, S. 3), die ja seit jeher zentralen Überprüfungen skeptisch gegenübersteht, sondern auch seitens der Bildungswissenschaft wurden erneut eine Reihe von Argumenten in die Diskussion eingebracht, die solche Testungen grundsätzlich zumindest kritisch hinterfragen (Meyer/Zahedi, 2014).

Klar scheint, dass Standardisierung und Testungen im Prinzip nichts Verwerfliches sind und auch im Sinne eines international-vergleichenden Bildungs-Monitorings durchaus ihre Meriten haben (können) und auch wertvolle Informationen über bestimmte Teilaspekte des Bildungswesens, insbesondere des Schulsystems, liefern und im Sinne einer „evidence-based“ Politik auch Systemsteuerung ermöglichen (Dobart, 2004, S. 561; Friedl-Lucyshyn/Benischek, 2010, S. 228). Evident ist gleichermaßen, dass in verschiedensten Ländern, wie Finnland, Frankreich, England u.a. (Gruber, 2004, S. 675; Ratzki, 2004, S. 680-682), zentrale Testungen auch im Sinne einer Berechtigungsfunktion versuchen, die Leistungsbeurteilung zu objektivieren und mehr Gerechtigkeit und Vergleichbarkeit zu schaffen (Altrichter/Schratz, 2004, S. 630; Specht/Freudenthaler, 2004, S. 619). Auch dies ist an sich nichts Neues oder per se Problematisches und ist wohl ein wertvoller Beitrag, mehr inter-institutionelle Gerechtigkeit ins Schulsystem zu bringen, sofern diesen Ansprüchen gerecht wird, was durchaus auch zu hinterfragen ist (von der Groeben, 2004, S. 687). Common Sense besteht in der Scientific Community allerdings auch darüber, dass in all den oben angesprochenen Testungen zwar ein relevantes, allerdings nur relativ kleines, Segment von Schule und Bildung oder von dem, was in der Schule gelernt werden sollte oder wird, testmäßig erfasst und evaluiert werden kann (Hameyer/Heckt, 2004, S. 8; Kraus, 2014, S. 15). Viele wichtige Bereiche und Kompetenzen werden, möglicherweise auch, weil der Aufwand oder die Kosten zu groß wären, gar nicht überprüft, und das führt natürlich zu einer gewissen Steuerung und Normierung in dem Sinne, dass jene Bereiche, die testrelevant sind, von den Lehrerinnen und Lehrern besonders in den Blick genommen werden (müssen) und andere „unterbelichtet“ bleiben. Es kommt dann nicht selten zu

Phänomenen wie „Teaching to the test“ (Gruber, 2004, S. 675; Messner, 2004, S. 702; Aigner, 2014) – auch als (negativer) Backwash-Effekt bezeichnet (Mayrhofer/Schallenberg, 2012, S. 46) – und ggf. zu einer „Vernachlässigung“ bestimmter Fächer bzw. Domänen (Regenbrecht, 2005, S. 18), was – wie noch zu zeigen sein wird – unter Umständen der Vielfalt im Unterricht nicht gerade zuträglich ist.

Problematisch wird es spätestens dann, wenn Testergebnisse dazu herangezogen werden, unangemessen zu generalisieren und gleich ganzen Schulsystemen aufgrund der Befunde, noch dazu aus wenigen Gegenständen bzw. Kompetenzbereichen, quasi „ein schlechtes Zeugnis“ ausgestellt wird (Liessmann, 2006, S. 76-77; Salcher, 2008, S. 202; Glattauer, 2011, S. 138). Noch bedenklicher ist es, wenn Verantwortungsträger/innen im Bildungsbereich (Politikerinnen und Politiker, Ministerien, Schulaufsicht, Direktorinnen und Direktoren usw.) aber auch Medien (Glattauer, 2011, S. 9) – bewusst oder unbewusst – die falschen Schlüsse aus den Ergebnissen ziehen und für ihre (strategischen) Ziele ge- oder noch schlimmer missbrauchen, etwa zur Argumentation für oder gegen die heiß umstrittene Gesamtschule (Helmke/Schrader, 2001, S. 82; Riegler, 2014).

Mich als eine noch immer in der Schule stehende Person betrifft es insofern, als ich einerseits nicht nur mit einer gewissen landauf landab grassierenden „Testitis“ konfrontiert werde – ständig wird der Unterricht durch die verschiedensten Evaluierungen und Testungen – meist von außen – „gestört“ oder zumindest unterbrochen. Andererseits sind mir schon Fälle bekannt, wo Kolleginnen und Kollegen von ihren Vorgesetzten (Direktorinnen und Direktoren, Schulaufsicht) durchaus hart – etwa mit den Ergebnissen der Bildungsstandards-Überprüfungen – in ihren Klassen konfrontiert wurden und sie Rechenschaft darüber ablegen mussten, warum „ihre“ Kinder im Vergleich zu anderen so „schlecht“ abgeschnitten haben. Nicht selten wird dann auch indirekt oder direkt, aber jedenfalls unzulässigerweise, ein Urteil über ihren Unterricht gefällt. Denn obwohl Bildungsstandards-Überprüfungen nicht dazu gedacht und konzipiert sind, die Leistungen von Lehrkräften zu überprüfen (Altrichter/Schratz, 2004, S. 643), sondern maximal, wie schon oben angemerkt, in gewissen Bereichen ein „System-Monitoring“ ermöglichen sollen, entsteht dadurch natürlich ein gewisser Druck auf Lehrerinnen und Lehrer, ihre Schülerinnen und Schüler dazu zu befähigen, (lediglich?) die Tests gut zu meistern, um sich Vorwürfe oder zumindest unangenehme Fragen zu ersparen. Das kann – wie schon gesagt – dazu führen, dass gewisse – testrelevante – Bereiche besonders unterrichtet und fokussiert werden (z.B. Sprache/n, Mathematik und eventuell Naturwissenschaften) und anderen (z.B. kreativen Gegenständen, Persönlichkeits- und Sozialkompetenzen) nicht der Stellenwert eingeräumt wird, den sie eigentlich verdient hätten (Gruber, 2004, S. 675). Ja selbst in testrelevanten Gegenständen kann das zu einer „Amputation“ des Faches bzw. zu einer „Verarmung“ des Unterrichtsgeschehens führen, wenn man sich vorwiegend an dem orientiert, was getestet wird (von der Groeben, 2004, S. 688-690; Mayerhofer/Schallenberg, 2012, S. 46).

Das Ziel dieses Aufsatzes soll es daher jedenfalls sein, einerseits die Möglichkeiten und Grenzen (inter)nationaler (zentraler) Testungen auszuloten und andererseits die Gefahren aufzuzeigen, die mit der Überbewertung dieser Überprüfungen verbunden sind. Insbesondere soll es auch darum gehen zu zeigen, dass es nicht nur nicht möglich ist, Bildung zu standardisieren oder gar (umfassend) zu messen, sondern dass mit diesen Bestrebungen sogar gefährliche Entwicklungen in Gang gesetzt werden, die dem umfassenden Bildungsauftrag – zumindest der österreichischen Schule – zuwiderlaufen. Letztlich setzen zentrale, standardisierte Testungen Normen, sie definieren quasi inhärent, was unter Bildung verstanden wird oder verstanden werden soll (Altrichter/Schratz, 2004, S. 642; Liessmann, 2006, S. 86), und das geschieht nicht immer in einem pädagogischen Sinn, sondern hat unter Umständen andere Gründe, die, wie manche Kritiker meinen, auch von einem neoliberalen, ökonomischen Zeitgeist getragen sind (Meyer/Zahedi, 2014). Deshalb soll dies bewusst ein skeptisch-kritischer Beitrag zu bildungspolitischen Trends sein, um auch Verantwortliche wachzurütteln und sie zumindest dazu anzuregen, ihr – möglicherweise unreflektiertes – Tun einer kritischen Prüfung zu unterziehen.

1.2 Fragestellungen

Der vorliegende Beitrag geht aufgrund der oben an- und ausgeführten Überlegungen von folgenden Forschungsfragen aus:

- Welche bedeutsamen nationalen und internationalen Testungen gibt es, und was sind ihre Zielsetzungen?
- Was können diese Testungen leisten, und wo liegen ihre Grenzen und auch mögliche Gefahren?

- Welche Konsequenzen könnten allenfalls aus den Studien und ihren Ergebnissen auf nationaler Ebene abgeleitet werden?
- Inwiefern fungieren diese Tests als Trendsetter und tragen zu einer Normierung und gegebenenfalls Einengung von Bildung und Unterricht bei?
- Inwiefern ist es überhaupt grundsätzlich möglich, Bildung zu standardisieren und zu (ver)messen?

1.3 Aufbau, methodische Vorgangsweise und Einschränkungen

Ausgehend von den unter 1.2 angeführten Fragestellungen sollen in einem ersten Schritt die bekanntesten nationalen und internationalen Studien und Testverfahren, wie PISA, PIRLS, TIMMS, Bildungsstandards und die teilzentrale standardisierte Reifeprüfung kurz vorgestellt und ihre Zielsetzungen analysiert werden. Im nächsten Abschnitt werden die Möglichkeiten und Grenzen dieser Testungen ausgelotet. Darauf sollen mögliche und unmögliche (!) Konsequenzen aus den Ergebnissen der diversen Überprüfungen abgeleitet werden, um im Abschnitt 2.4 die Gefahren der implizit und manchmal sogar explizit vorgenommenen Standardisierung und Normierung von Bildung aufzuzeigen. Abschließend soll in diesem Sinne auch die Unmöglichkeit der Messung von Bildung generell herausgearbeitet werden.

Die oben vorgestellten Fragestellungen werden auf hermeneutischem Weg mittels Sekundäranalyse relevanter Literatur bearbeitet.

Klargestellt soll sein, dass ich im Sinne eines konstruktiv-kritischen Zugangs dieses Artikels bewusst einen Kontrapunkt zur allgegenwärtigen Standardisierungseuphorie setzen möchte. In dieser Hinsicht ist er aus Sicht mancher Befürworterinnen und Befürworter der nationalen und internationalen Studien und Tests unter Umständen wissenschaftlich nicht ausgewogen bzw. objektiv genug. Da absolute Objektivität aus meiner Sicht aber ohnehin nicht existiert und immer subjektiv bleibt, schon aufgrund der „Störvariable Mensch“ in bildungswissenschaftlichen u.a. Kontexten, bezieht die Arbeit aber ganz klar Stellung und lädt ein zum Diskurs.

Aufgrund der gebotenen Kürze des vorliegenden Beitrags können verschiedene Themen und Argumente nur angerissen und im Überblick diskutiert werden, was möglicherweise eine (weitere) Einschränkung darstellt. Der Verweis auf die einschlägige Literatur soll aber den interessierten Leser oder die interessierte Leserin ermutigen, sich eingehender mit der Problematik auseinander zu setzen. Diese Arbeit soll jedenfalls nicht in irgendeiner Schublade „verschwinden“, sondern sie soll Ausgangspunkt und Anlass für Interessierte sein, sich in der Thematik noch zu vertiefen und die Ergebnisse auf Basis einer „critical friendship“ auch den bildungspolitisch Verantwortlichen und Entscheidungsträgerinnen bzw. -trägern im Sinne eines konstruktiven Feedbacks zur Verfügung zu stellen.

2 Was uns nationale und internationale Leistungsvergleiche erzählen und was nicht

2.1 PISA, PIRLS, TIMSS, Bildungsstandards, „Zentralmatura“ & Co – oder: Ist die „Testitis“ ausgebrochen? Eine Einführung in den „Testwahnsinn“

Das Thema „internationale und nationale Schulleistungsvergleiche“ muss man hinsichtlich seiner Genese im Längs- und im Querschnitt im historischen und internationalen Kontext sehen. Dabei fällt vor allem auf, dass es gewisse „Wellenbewegungen“ gegeben hat, die einmal mehr den Zentralismus im Schulwesen, einmal mehr die Autonomie im Blick hatten. Es ist außerdem festzustellen, dass internationale Trends nicht gleichzeitig auftraten und -treten, sondern auch immer mit einer gewissen Zeitverzögerung in einzelnen Ländern und Erdteilen (nach)vollzogen wurden und werden. Auch die weltweite und insbesondere europäische Integration trug das Ihre dazu bei. An zwei Beispielen soll das exemplarisch festgemacht werden: England und Österreich. Während z.B. in England bis in die 1980er Jahre die sogenannten „Local Educational Authorities“ (LEAs) – also lokale Schulbehörden – das Sagen hatten und das Schulsystem extrem dezentralisiert war (Gillard, 2011), gab es hierzulande damals ein sehr stark vom Bildungsministerium gelenktes und starres Bildungswesen, das kaum

Autonomie für einzelne Schulen kannte (Lentner, 2008, S. 22), oder um es mit Vierlinger (1993, S. 71) auszudrücken: „Die österreichische Schulreform der siebziger Jahre war eine Demonstration der Stärke zentraler Steuerung.“

Die große Autonomie führte in England zu einem „blühenden“, bunten Schulsystem, bei dem der einzelne Schulstandort sehr viel Gestaltungsmöglichkeiten und sogar eine relativ große Budgethoheit hatte. So konnten etwa die Mittel und Ressourcen sehr effektiv und nach dem (lokalen) Bedarf eingesetzt werden. Ob man Infrastrukturanschaffungen machte oder mehr Lehrerinnen und Lehrer beschäftigte, und vor allem welche, das oblag im Wesentlichen den LEAs bzw. der einzelnen Schule (Vierlinger, 1993, S. 89-90). Der Nachteil daran war natürlich einerseits, dass ein Schulwechsel für Kinder nicht immer einfach war, da es mitunter an der neuen Schule ganz andere Curricula und Schwerpunkte gab, und andererseits, dass Leistungen kaum bis gar nicht vergleichbar waren (Vierlinger, 1993, S. 107; Gillard, 2011).

In Österreich, im Gegensatz dazu, wurden etwa die Mittelzuteilung und auch die curricularen Vorgaben sehr restriktiv zentral organisiert. Auch die Personalhoheit lag (und liegt bis heute) nicht an den einzelnen Schulen, was dazu führte, dass man sich am Standort leicht auf „die da oben“ ausreden konnte, wenn etwas nicht funktionierte, und außerdem sowohl Ergebnisverantwortlichkeit als auch -bewusstsein fehlte (Vierlinger, 1993, S. 71).

Interessanterweise schlug das Pendel in beiden Ländern zu Beginn der 90er Jahre des vorigen Jahrhunderts – aus verschiedenen Gründen – genau in die entgegengesetzte Richtung aus: In England erkannte man – auch aufgrund einer aufkeimenden Testkultur – die recht unterschiedlichen Leistungen und Standards bzw. Niveaus an den einzelnen Schulen, und auch vor dem Hintergrund der Globalisierung und Europäisierung erschien dieses dezentrale Schulsystem nicht mehr zeitgemäß (Vierlinger, 1993, S. 108; Gillard, 2011). In Österreich erkannte man, dass die Bedürfnisse in den einzelnen Regionen sehr unterschiedlich sind, etwa in Ballungsräumen ganz anders als im ländlichen Raum, und so war eine Steuerung mit der „Rasenmähermethode“, also „allen das Gleiche“, kaum noch zu rechtfertigen, vor allem, da auch die Klagen von der Basis (Schulen, Lehrkräfte) immer lauter wurden (BMUK, 1993, S. 3). Außerdem wurden die Bildungsbudgets immer knapper, und der Verdacht besteht, dass das ein nicht unerheblicher Grund war, warum man sich vor etwa 20 Jahren dazu entschied, den Schulen mehr Autonomie zu gewähren (Forum Politische Bildung, 2014). Denn dann waren ja die Schulen verantwortlich, mit den – nun verringerten – Budgets auszukommen. Das war die Zeit, als etwa Musik- und bilinguale Volksschulen, Öko-, Sport- und Informatikhauptschulen entstanden (BMUK, 1993, S. 2). Den Schulen wurde also ein gewisser curriculärer Spielraum gegeben, etwa mittels Schwerpunktsetzung in einigen Gegenständen, um ein eigenes Schulprofil zu entwickeln – natürlich in mehr oder weniger eingeschränktem Rahmen. Über wirkliche (satte) Budget- oder gar Personalhoheit verfügten die Direktorinnen und Direktoren natürlich nach wie vor nicht (BMUK, 1993, S. 8; Lentner, 2008, S. 28). Auch angesichts der immer stärker werdenden Verlagerung der Schülerströme von den Hauptschulen zu den Gymnasien erschien das ein probates Mittel, um die Hauptschulen attraktiver zu machen und dieses „Creaming“ zu verhindern oder zumindest zu verlangsamen. Da aber auch die Allgemeinbildenden Höheren Schulen die angesprochenen schulautonomen Spielräume bekamen, setzte ein richtiger Wettkampf um Schüler(zahlen) ein, auch weil es einen markanten Schülerrückgang (aufgrund der sinkenden Geburtenrate) gab. Und dieser Trend, der oft zu Ungunsten der Hauptschulen und nunmehr Neuen Mittelschulen ausging und -geht, hält bis heute ungebrochen an (Lentner, 2008, S. 22-24).

Diese Autonomie hatte natürlich auch einige unangenehme Folgen: Zum einen war an den Standorten entscheidend, welche (Fächer-)Lobbys das Sagen hatten und dementsprechend gingen natürlich auch die Kürzungen in einzelnen Gegenständen zugunsten anderer aus (Lentner, 2008, S. 103). Der Verfasser dieser Zeilen erlebte, als Lehrer der Gegenstände Physik und Chemie, selbst leidvoll mit, dass die naturwissenschaftlichen Fächer, ausgenommen in den entsprechenden Schwerpunktschulen, an vielen Standorten stark reduziert wurden. Der Gegenstand Physik/Chemie, damals in der Hauptschule auf der achten Schulstufe ein Vier-Stunden-Fach, wurde z.B. vielerorts auf zwei Stunden gesenkt, um andere Gegenstände bzw. Schwerpunkte quasi „aufzufetten“, oder gar neue Fächer zu schaffen, etwa das Fach „Soziales Lernen“ – als ob sich das überhaupt veranstalten lässt. So kam es dazu, dass gewisse Bereiche des Lehrplans nur mehr basal abgedeckt wurden und werden und dass wohl zu Recht ein Aufschrei aus Naturwissenschaft und Technik erfolgte.

Die schulautonomen Profile führten in weiterer Folge natürlich dazu (wie schon in England), dass die Niveaus in einzelnen Fächern (je nach Schwerpunktsetzung) an den einzelnen Schulen auseinander drifteten, was ein natürlicher Nebeneffekt der Freiräume war. Offenbar war das dem Bildungsministerium dann doch nicht ganz geheuer (erlauben Sie mir diese subjektive Wertung), weshalb man einen Steuerungsmechanismus suchte, der gewisse Niveaus in den einzelnen Fächern garantierte. In den Bildungsstandards schien man ihn

gefunden zu haben (Dobart, 2004, S. 560-561). Den Schulen wurden nun zu erreichende Standards in bestimmten ausgewählten Kompetenzbereichen gesetzt (mittlerweile bereits auf gesetzlicher Basis), der Weg dorthin blieb den Standorten überlassen. Natürlich forderte man auch eine gewisse Ergebnisverantwortlichkeit (im Sinne einer Output-Orientierung) ein, was natürlich den Druck auf die Direktionen und Lehrkräfte erhöhte. Dazu kam und kommt noch, dass die Schulen ja kaum wirklich solche Spielräume hatten und haben, wie sie nötig gewesen wären oder sind, um tatsächlich autonom effektiv zu handeln. Ziel war es jedenfalls, ein gewisses Basiswissen in einzelnen Fächern zu garantieren bzw. sicher zu stellen und eine Vergleichbarkeit der Niveaus und Inhalte zu gewährleisten und diese auch zu überprüfen (Dobart, 2004, S. 560; Forum Politische Bildung, 2014).

Mit der Jahrtausendwende wurden deshalb in einzelnen Gegenständen (Deutsch und Mathematik auf der vierten Schulstufe; Deutsch, Mathematik und Lebende Fremdsprache auf der achten Schulstufe) Kompetenzen und in weiterer Folge Standards formuliert, die die Schüler/innen erreichen soll(t)en (Kern, 2004, S. 566-569; Rieß/Meließnig, 2010, S. 358-360). Der lange (semantische und auch inhaltliche) Streit darüber, ob das Mindest-, Regel-, oder Orientierungsstandards sein sollten, kann hier aus Platzgründen gar nicht erörtert werden (Altrichter/Schratz, 2004, S. 634-635). Gruber (2004, S. 671) formulierte es wie folgt: „Es entbehrt nicht einer gewissen Ironie, dass ausgerechnet der Begriff `Standards` (...) in einer Fülle unterschiedlicher Bedeutungen auftaucht und interpretationsbedürftig ist wie kaum ein anderer pädagogischer Terminus technicus“. Schließlich einigte man sich auf Bildungsstandards, ein denkbar unglücklicher und unangemessener, weil völlig überzogener Ausdruck, wie noch zu zeigen sein wird. Bildungsstandards sollen gewisse nachhaltige, fachliche Grundkompetenzen beschreiben, die von – den meisten – der Schülerinnen und Schüler der jeweiligen Altersstufe erreicht werden. Sie basieren auf den Lehrplänen und dienen von ihrer Konstruktion und Intention her nicht der Leistungs- oder gar Lehrerbeurteilung, sondern sollen einerseits den Lehrkräften Orientierung bieten, was in den ausgewählten Gegenständen unter „normalen“ Umständen erreicht werden kann und soll, und andererseits sind sie – wie schon gesagt – ein Instrument, das System-Monitoring und Evaluierung erlauben soll. So können sie zur Qualitätssicherung und -entwicklung ebenso verwendet werden wie als Anregung und Feedback für Lehrerinnen und Lehrer – was unterrichtet werden kann/soll und wo sie leistungsmäßig mit ihren Klassen stehen (Beer, 2005, S. 478-479). Der damalige Sektionsleiter im Bildungsministerium, Anton Dobart, formulierte in diesem Zusammenhang ein hehres Ziel der Bildungsstandards: Sie sollen dazu dienen „(...) dass Schülerinnen und Schüler besser lernen, Lehrerinnen und Lehrer besser unterrichten und Schulen effektiver arbeiten können“ (Dobart, 2004, S. 562). Natürlich bieten Bildungsstandards auch die Möglichkeit, Schulen und Schulbezirke bzw. auch Bundesländer genauso wie Schularten miteinander zu vergleichen, was hinsichtlich der oft diskutierten Frage der „Systemgerechtigkeit“ interessant erscheint (Beer, 2005, S. 480-481). Ein Manko ist natürlich, dass Bildungsstandards, wie auch PISA & Co, keine Auskunft darüber geben, unter welchen Bedingungen gewisse Leistungen zustande gekommen sind oder kommen. Wagner (2010, S. 72) formulierte es ein wenig sarkastisch: „Was passiert, wenn angenommen eine Brennpunktschule schlecht abschneidet? Wird den Schülern noch einmal bestätigt, dass sie leistungsschwach sind und die `Loser` der Nation? Werden die Lehrer, die unter schwierigen Bedingungen unterrichten mußten¹, als Versager gebrandmarkt? Oder kommt der Großinquisitor vom real existierenden BIFIE selbst, um allen zu zeigen, wie Schule geht?“ Das hat schon was: So kann ein mittelmäßiges Ergebnis einer Klasse in dem einen Fall ein großartiger Erfolg sein (da etwa die Voraussetzungen der Klasse, wie etwa der Anteil von Kindern, die die Unterrichtssprache nicht verstehen, eher nachteilig waren), in dem anderen Fall kann es eher ein Misserfolg sein (wenn die Rahmenbedingungen, z.B. Ausstattung, Leistungsfähigkeit der Kinder, günstig waren). Die Interpretation der Ergebnisse muss deshalb immer vor Ort erfolgen und bedarf einer durchaus reflexiven und zumindest auch basalen wissenschaftlichen Kompetenz. Gerade deshalb wurden und werden auch „Rückmeldemoderatorinnen und -moderatoren“ ausgebildet, die den Verantwortlichen helfen sollen, die richtigen Schlüsse aus den Ergebnissen zu ziehen (Friedl-Lucyshyn/Benischek, 2004, S. 231-232). Denn gerade in diesem Bereich kann eine inkompetente Deutung der Daten durchaus Schaden anrichten und der ganzen Sache keinen guten Dienst erweisen. Man wird sehen, wie sich dieses Projekt weiterentwickelt und ob Bildungsstandards tatsächlich „nur“ jenen Stellenwert bekommen, den sie auch verdienen. Denn eines ist klar: Überschätzen darf man ihre Bedeutung keinesfalls, wie noch zu zeigen sein wird.

Die letzte „Errungenschaft“ und auch quasi logische Konsequenz der Befürworterinnen und Befürworter zentraler Testungen ist in Österreich die standardisierte, kompetenzorientierte Reife- bzw. Diplomprüfung – umgangssprachlich und fälschlicherweise auch als „Zentralmatura“ bezeichnet, da sie bestenfalls teils zentral ist; denn nur in einigen schriftlichen Gegenständen werden die Aufgaben zentral seitens des BIFIE vorgegeben.

Schon die (mehr oder weniger) standardisierte Auswertung erfolgt an den Standorten und die mündliche Reifeprüfung ist ohnehin – nach wie vor – Sache der einzelnen Schulen bzw. Lehrerteams. Ab dem Schuljahr

2015/16 wird diese neue Form der „Matura“ erstmals bundesweit (zumindest an den Allgemeinbildenden Schulen) durchgeführt, und es bleibt abzuwarten ob – trotz der anfänglichen Pannen des BIFIE – diese Innovation das hält, was sich ihre Initiatorinnen und Initiatoren versprechen (Braun/Weidinger, 2012). Und damit wieder zurück auf die internationale Bühne.

In England ging man Anfang der 1990er Jahre einen noch rigideren Weg: War bis dahin, wie erörtert, die Schullandschaft sehr dezentral organisiert, schüttete man das Kind dann völlig mit dem Bade aus. Man entmachtete die LEAs und schuf ein extrem zentralistisches Schulsystem mit detaillierten Curricula, mit präzisen Vorgaben bis (hinunter) auf Lernzielebene, und die Schulen – oder genauer gesagt die Lehrkräfte – hatten und haben sich daran fast sklavisch zu halten, insbesondere, weil man an bestimmten Key Stages des Schulsystems nationale Testungen einführt, die allerdings – anders als die Bildungsstandards – auch für die Beurteilung von Schülerinnen und Schülern herangezogen wurden (Vierlinger, 1993, S. 94). Das schuf natürlich auf der einen Seite eine extreme Vergleichbarkeit der Leistungen, führte aber auf der anderen auch zu einem richtigen „Rankingwahn“ (Ratzki, 2004, S. 681). Auch die Lehrerrolle veränderte sich schlagartig vom Beurteiler bzw. der Beurteilerin zum Coach. Der pädagogisch-didaktischen und auch der humanen Sache tat man dabei allerdings keinen besonderen Gefallen. Dies führte nämlich gerade an den Key Stages zu einem regelrechten „Teaching to the test“, was zwangsläufig den Druck auf Schülerinnen und Schüler, wie auch Lehrkräfte gleichermaßen, erhöhte und mit Sicherheit zu einer inhaltlichen und didaktischen Einengung und wohl auch „Monokultur“ des Unterrichts führte (Vierlinger, 1993, S. 118; Ratzki, 2004, S. 681; Gillard, 2011) – eine Gefahr, die zwar im geringeren Ausmaß, aber doch, auch in Österreich besteht, und generell überall dort auftritt, wo zentrale Testungen Berechtigungen verleihen und so über das Wohl und Wehe von Schülerinnen und Schülern (oft aber auch Lehrerinnen und Lehrern) entscheiden.

Parallel zu diesen Entwicklungen in England und Österreich setzte international auch die Ära der großen Bildungsstudien ein, die in weiterer Folge die politischen Verantwortungsträger/innen gleichsam vor sich her trieben – speziell in jenen Ländern, die im internationalen Vergleich schlecht abschnitten (Messner, 2004, S. 695): Die „Mutter“ all jener Studien war etwa die TIMSS-Studie (Trends in International Mathematics and Science Study), nach wie vor die größte internationale Vergleichsstudie. Erstmals 1995 durchgeführt, misst sie im Vier-Jahres-Rhythmus Mathematik- und Naturwissenschaftskenntnisse auf der vierten Schulstufe (Bifie, 2014g). 2001 folgten gleich zwei weitere bekannte Studien: PIRLS (Progress in International Reading and Literacy Study) testet Schülerinnen und Schüler der vierten Schulstufe im Fünfjahres-Zyklus in der Lesekompetenz (Bifie, 2014c). Und die berühmteste von allen, PISA (Programme for International Student Assessment), fokussierte erstmals auf Schülerinnen und Schüler der Sekundarstufe: 15- bis 16-Jährige werden seit 2001 alle drei Jahre in den Bereichen Lesen, Mathematik und Naturwissenschaften überprüft (Bifie, 2014e).

Es ist sattsam bekannt, dass Österreich in den diversen Studien seit damals im Durchschnitt bestenfalls mittelmäßig abschneidet, während Finnland oder fernöstliche Länder, um nur einige Beispiele zu nennen, regelmäßig an der Spitze der Testungen liegen (Glattauer, 2011, S. 63; Bifie, 2013). Doch um die Ergebnisse im Einzelnen soll es an dieser Stelle nicht gehen – das würde zu weit führen; außerdem müsste man dann auch viel genauer die Unterschiede innerhalb eines Landes analysieren. Die österreichischen Gymnasien schneiden z.B. im internationalen Vergleich sehr gut ab. Viel eher soll in weiterer Folge erörtert werden, was solche Studien generell leisten können, welche Schwachstellen sie haben und welche Gefahren damit unter Umständen verbunden sein könnten.

2.2 Vom Sinn und Unsinn nationaler und internationaler Testungen – oder: Chancen, Risiken, Möglichkeiten und Grenzen

Die vorliegende Arbeit soll keinesfalls den Eindruck erwecken, dass sie zentrale nationale und internationale Studien und Bildungsvergleiche grundsätzlich ablehnt. Natürlich haben diese einen gewissen Stellenwert, der nicht geringgeschätzt werden darf. So liefern sie – wie schon oben ausgeführt – allesamt etwa wertvolle Daten über Bildungssysteme bzw. – genauer gesagt – Ausschnitten daraus und ermöglichen in Teilbereichen durchaus Leistungsvergleiche. Selbstverständlich bekommen dann einzelne Länder Hinweise darauf, wie es um die Leistungen ihrer Schülerinnen und Schüler in bestimmten Kompetenzbereichen steht. So haben wir in Österreich aufgrund der internationalen Studien – aber auch nationaler – etwa erfahren, dass es unser Schulsystem nur relativ schlecht schafft, sowohl die guten als auch die schlechten Schülerinnen und Schüler optimal zu fördern. Ca. ein Fünftel der Getesteten schneidet hervorragend ab, jedoch jeder bzw. jede Dritte ist ein so genannter „Risikoschüler“ bzw. eine „Risikoschülerin“, etwa im Bereich Lesen (z.B. Bifie, 2014d). Andere Schulsysteme schaffen es, auch unter vergleichbaren Bedingungen, besser, die Kluft zwischen den Spitzen- und

Minderleistungen deutlich enger zu halten, ohne dass das auf Kosten der sehr guten Schülerinnen und Schüler geht. Ein weiterer Hinweis ergibt sich aus der Tatsache, dass die Kinder mit Migrationshintergrund in unserem Bildungssystem signifikant zu den „Verlierern“ zählen (Bruneforth/Herzog-Punzenberger/Lassnig, 2012, S. 12), mehr jedenfalls als in anderen Ländern. Ein Beispiel: „In der Lese- und Mathe-Risikogruppe macht der Anteil an PISA-Migranten 28 Prozent aus, in der Naturwissenschaftsgruppe sogar 34 Prozent – und das bei einem Anteil von nur 15 Prozent an der Gesamtgruppe“ (Glattauer, 2011, S. 26). Kritisch angemerkt wird dazu auch immer wieder, dass in Österreich, wie in Deutschland, nach wie vor Bildung „vererbt“ ist bzw. wird: In nur wenigen Ländern hängt der Testserfolg bei PISA & Co offensichtlich so stark vom Bildungshintergrund der Eltern ab wie hierzulande oder bei unseren bundesdeutschen Nachbarn (Bacher, 2008, S. 20; Wößmann, 2008, S. 513; Bifie, 2014d). Unser System schafft es also scheinbar nur sehr schlecht, diese Eingangsdefizite der Schülerinnen und Schüler, etwa aus bildungsfernen Schichten, auszugleichen. Dieses Argument wird immer wieder ins Treffen geführt, insbesondere wenn es darum geht, auf die „Meriten“ der Gesamtschule hinzuweisen. Ignoriert wird dabei oft, dass die „Phrase von der vererbten Bildung“ auf ungenauen Datensätzen der OECD (über Bildungsabschlüsse) beruht, wie Feller (2015) nachweisen konnte. Eine detailliertere Auswertung ergab, dass unser Bildungssystem gar nicht so sozial undurchlässig ist, wie vielfach behauptet wird.

Ein Vorwurf, der dem österreichischen Schulsystem immer wieder gemacht wird, ist auch jener, dass der (budgetäre) Aufwand durchaus nicht die Ergebnisse bringt, die man erwarten könnte. Österreich leistet sich – kritischen Stimmen zufolge – ein relativ teures Schulsystem (Salcher, 2008, S. 195-196; Glattauer, 2011, S. 177), bei dem die Ergebnisse mehr oder weniger um den OECD-Durchschnitt schwanken. Dem darf aber nicht vorbehaltlos zugestimmt werden. Denn das in dieser Hinsicht eher unverdächtige Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des österreichischen Schulwesens kommt im Nationalen Bildungsbericht 2015 zu folgendem Schluss: „Für den Primar- und Sekundarbereich zusammen gibt Österreich 6,1 % der Staatsausgaben aus, das ist der viertniedrigste Wert aller OECD-Länder“ (Bruneforth u.a., 2016, S. 325). Dabei liegt es natürlich nicht nur am Geld, sondern auch an dessen effektivem Einsatz. Denn es zeigte sich schon im Rahmen der McKinsey-Studie, dass Geld allein noch kein gutes Schulsystem garantiert (Barber/Mourshed, 2007). Finnland gibt z.B. annähernd denselben Prozentsatz des BIPs für Bildung und Schule aus (6,8 %, 2015), jedoch kommt dort viel mehr Geld direkt in der Klasse an, etwa in Form von Unterstützungssystemen, von denen Lehrerinnen und Lehrer in unserem Land nur träumen (von Freymann, 2002). Bei uns „versickert“ viel des aufgewendeten Geldes in der Verwaltung, in „Doppelgleisigkeiten“ und Mehrfachzuständigkeiten, wie schon mehrere, auch nationale, Studien belegten (IHS, 2007). Und in diesem Zusammenhang gewinnt die langjährige Forderung nach einer längst überfälligen Verwaltungsreform – speziell im Bildungsbereich – neue Bedeutung. Ob diese politisch wirklich gewollt und durchführbar ist, bleibt abzuwarten; die Vorzeichen stehen angesichts der immer wieder aufflammenden und ungelösten Diskussion, ob die Zuständigkeit für Schule und Bildung eher Bundes- oder Landessache sein soll, eher negativ. Die internationalen Studien legen also durchaus den Finger in die Wunden, die bis Mitte der 1990er Jahre aufgrund fehlender Daten gar nicht bekannt waren. Insofern können solche Befunde nur ein Gewinn sein, wenn man sie kompetent und nicht einseitig deutet.

Nationale Testungen, wie Bildungsstandards oder die „Zentralmatura“, haben natürlich ganz andere Stoßrichtungen. Während die erstgenannten lediglich evaluativen Charakter haben, verleiht die standardisierte, kompetenzorientierte, teilzentrale Reifeprüfung Berechtigungen, und es wird der Versuch unternommen, die Leistungen der Absolventinnen und Absolventen, zumindest in Teilbereichen, zu standardisieren bzw. vergleichbar zu machen (Schatzl, 2012, S. 9-10). Das kann natürlich zu einer gewissen Bildungsgerechtigkeit führen, zumindest aber zu einer Vergleichbarkeit, die nicht von Nachteil ist. Zweifellos werden die Ergebnisse allerdings auch vor Augen führen, wie unterschiedlich die Niveaus an den einzelnen Standorten unter Umständen tatsächlich sind und dass für gleiche Leistungen oft unterschiedliche – oder keine – Berechtigungen verliehen werden, oder anders gesagt: dass es hinsichtlich des Erfolgs bei der Matura bislang darauf angekommen ist, in welcher Schule oder bei welcher Lehrkraft man „sitzt“ – ein Phänomen, das wir von der (subjektiven) Leistungsbeurteilung, die den basalen Gütekriterien, wie Objektivität, Reliabilität, Validität oder Vergleichbarkeit nicht genügt, ohnehin von allen Schulstufen und aus allen Schularten seit Langem kennen (Ingenkamp, 1971; Vierlinger, 1993, S. 51-54).

Keinesfalls eignen sich solche Studien allerdings dafür, Aussagen über den Erfolg oder Misserfolg von Bildungssystemen insgesamt zu treffen, dafür sind sie viel zu schmal angelegt und messen maximal einen Bruchteil dessen, was unterrichtet wird und auch werden soll. Eine große Gefahr besteht z.B. auch darin, dass die Ergebnisse einerseits mehr versprechen als sie halten können und andererseits, dass sie als Anhaltspunkt für schulpolitische Bildungsentscheidungen eines Landes herangezogen werden (Liessmann, 2006, S. 76-77; Glattauer, 2011, S. 138). Salcher (2010) formulierte es für die größte internationale Leistungsvergleichsstudie

sehr pointiert: „PISA ist tatsächlich nicht mehr als eine klassische Schularbeit. (...) PISA misst die Geschwindigkeit des Fließbands und bewertet die Präzision der Massenfertigung eines völlig veralteten Schulsystems, das auf dem Modell des 19. Jahrhunderts aufbaut“ (Salcher, 2008, S. 202). Oder um ein altes Sprichwort zynisch abzuwandeln: „Nicht für die Schule, nicht für das Leben, für die Standardtests lernen wir. (...) Standardtests stellen lediglich eine Kompetenz zweifelsfrei fest: Ob und in welchem Umfang ein Proband den Standardtest schafft, also die Standardteststauglichkeit“ (Wagner, 2010, S. 70). Gerade die PISA-Studie hat aber in vielen Ländern dazu geführt, dass „OECD-hörige“ Politikerinnen und Politiker das ganze Schulsystem aufgrund der Ergebnisse „umkrepeln“ woll(t)en. In Deutschland und Österreichs, wo man fast einen „PISA-Schock“ angesichts der recht durchwachsenen Ergebnisse erlebte, orientierten sich viele Bildungspolitiker/innen in ihren Überlegungen an den angeblich erfolgreichen Ländern, insbesondere Finnland. Heerscharen von Verantwortungsträgerinnen und -trägern pilgerten in diese Länder und überlegten, wie man die scheinbar erfolgreichen Systeme oder zumindest Teile davon in die Heimatländer übertragen (quasi „importieren“) könnte. Dabei vergaßen und vergessen sie aber zum einen den Umstand, dass die PISA-Tests nichts über die Qualität eines Bildungssystems aussagen, und zum anderen, dass man Äpfel mit Birnen nicht gut vergleichen kann. Vieles, was in Finnland und anderen „PISA-Spitzenländern“ zum Erfolg beiträgt, ist schlicht und ergreifend nicht importierbar, da Schulsysteme nicht nur historisch gewachsen, sondern auch eingebettet in eine Kultur und Tradition eines Landes oder einer geopolitischen Region sind, die mit der heimischen einfach nicht zu vergleichen sind (Preisendörfer, 2008, S. -176). So wäre es etwa völlig kontraproduktiv, das durchaus testmäßig erfolgreiche japanische Schulsystem auf Österreich zu übertragen, denn das hieße die gesamte dort herrschende Denkweise, Kultur und Philosophie wie auch Bildungsaspiration mit zu „übersiedeln“ – und das wünschen wir uns in Kenntnis der Verhältnisse in Japan speziell und in Asien insgesamt für unsere Länder wohl eher nicht (Glattauer, 2011, S. 62-75). Selbst finnische Verhältnisse wären nicht herstellbar, da auch dort gewisse gesellschaftliche Voraussetzungen völlig unterschiedlich zu den unseren sind (von Freymann, 2002); aber gerade die soziokulturellen Bedingungen sind es, die das gute oder schlechte Abschneiden bei PISA und anderen internationalen Studien zu einem nicht unwesentlichen Teil beeinflussen und erklären. Explizieren wir das am Beispiel Finnlands ein wenig:

Finnland hat eine sehr homogene Gesellschaft und Bevölkerungsstruktur (wie etwa die meisten asiatischen Staaten auch) mit nur ca. 2-4 % Migrationsanteil, was es im Vergleich mit Österreich natürlich leicht macht, Kinder mit nichtfinnischer Muttersprache entsprechend zu integrieren und schulisch (etwa in der Unterrichtssprache) zu fördern.

Weiters findet man in Finnland eine lange Lesetradition, vermutlich aufgrund der langen, kalten Nächte und Winter; die Bibliotheken sind kostenlos und Bibliotheksbusse fahren regelmäßig in den hohen Norden. Und da Lesen-können der Schlüssel zu einem erfolgreichen Lernen ist, sind die Finninnen und Finnen hier sicher im Vorteil.

Ausländische TV-Programme, Serien und Filme werden – aufgrund des geringen „Zusehermarktes“ – meist nicht synchronisiert und in der englischen Originalversion inklusive Untertitel ausgestrahlt. Das hat einen zweifachen Vorteil: Zum einen gewöhnen sich die Kinder sehr schnell und unbewusst an die englische Sprache und erlernen diese deshalb auch vergleichsweise leicht, zum anderen haben sie schon früh eine hohe Motivation, lesen zu können, um die Untertitel zu verstehen.

Nicht zu unterschätzen ist die Tatsache, dass das Finnisch eine phonetische Orthografie aufweist, was es relativ leicht macht, Lesen und Schreiben zu erlernen; es ist in diesem Zusammenhang kein Zufall, dass Finnland weltweit die niedrigste Analphabetenrate (ca. 4 %) aufweist (von Freymann, 2002).

Diese und andere Faktoren und Rahmenbedingungen kann man einfach nicht nach Österreich „importieren“ und sie führen einfach dazu, dass schon die Bedingungen des Aufwachsens und Lernens in beiden Ländern einfach zu unterschiedlich sind. Und sie erklären eben, wie schon gesagt, einen relativ hohen Anteil an den Testergebnissen.

Andererseits darf man aber auch nicht übersehen, dass etwa die Jugendarbeitslosigkeit in Finnland deutlich über der österreichischen liegt (Mohr, 2012) und auch andere Indikatoren (z.B. Schulzufriedenheit, Selbstmord- und Alkoholikerrate) eher unserer Gesellschaft ein gutes Zeugnis ausstellen (Roth, 2014). Mit einem Wort: Soziokulturelle Bedingungen sind einfach nicht vergleichbar und der „Erfolg“ eines Schulsystems lässt sich möglicherweise auch an anderen Kriterien und Indikatoren ablesen als an den PISA-Ergebnissen. Auch eignen sich Studien wie PISA kaum, etwa pro oder contra Gesamtschule zu argumentieren. Der Hinweis, dass nahezu alle „PISA-Sieger“ gesamtschulisch organisiert sind, vernachlässigt die Tatsache, dass sich auch unter den PISA-Verlierern (in den einzelnen Gegenständen) fast nur Gesamtschulländer befinden (Bifie, 2013) – schlicht und ergreifend deshalb, weil diese Art der Organisation von Schulsystemen eben international viel mehr die Regel denn die Ausnahme darstellt (Wößmann, 2008, S. 510). Andererseits sei auch darauf hingewiesen, dass es DIE

Gesamtschule nicht gibt, sondern dass sich die Organisationsformen hier sehr stark unterscheiden und z.T. sehr viel – auch äußere – Differenzierung stattfindet. Stefan Hopmann, ein intimer Kenner nationaler und internationaler Schulsysteme, meinte dazu einmal sehr klar: „Die erhoffte soziale Chancengleichheit ist ein Irrtum. Fast alle Gesamtschulländer haben alternative Formen der Segregation entwickelt“ (Hopmann, 2016, zit. nach Hartweger, 2016).

2.3 „Nur weil man die Kuh öfter wiegt, gibt sie dennoch nicht bessere und mehr Milch“ – oder: Mögliche und notwendige Konsequenzen aus den Testergebnissen

Nationale und internationale Testungen liefern, wie bereits ausgeführt, durchaus wertvolle Informationen über den Leistungsstand von Schülerinnen und Schülern in ausgewählten Teilbereichen des Schulsystems. Diese Daten können natürlich zum Zwecke der Bildungssteuerung verwendet werden, sofern sie frei von (reinen) parteipolitischen, weltanschaulichen Intentionen sind und mit aller gebotenen Vorsicht interpretiert werden. Das war – wie die Erfahrung gezeigt hat – leider nicht immer der Fall, sehr auch zum Unmut von Bildungsexpertinnen und -experten (insbesondere Bildungswissenschaftlerinnen und Bildungswissenschaftlern), die – anders als Politik und Medien – diese neue entstandene Testkultur durchaus differenziert und kritisch sehen und auch analysieren.

Es gilt also, aus den Testergebnissen die richtigen und angemessenen Schlüsse zu ziehen und nicht – vorschnell – politische Forderungen für ein ganzes Bildungssystem abzuleiten (etwa die Gesamtschule einzuführen). Denn natürlich haben die Ergebnisse von PISA & Co auch einige wichtige Impulse gesetzt, etwa dahingehend, dass Bildung plötzlich zu einem heiß diskutierten Thema wurde (Glattauer, 2010, S. 14; Glattauer 2011, S. 10), dass gewisse Aspekte in den Blickpunkt der Öffentlichkeit rückten, die bislang unbekannt, verschleiert und auch möglicherweise bewusst unter (politischem) Verschluss gehalten wurden. Diese Ergebnisse bieten deshalb natürlich auch Lernchancen für die Verantwortlichen.

Aber: Es entsteht alles in allem der Eindruck, dass zwar die Probleme, die beispielsweise unser Schulsystem aufweist (wie oben schon diskutiert), hinlänglich erforscht und bekannt sind, die Maßnahmen und Konsequenzen jedoch – möglicherweise aus budgetären Gründen – nicht in dem Ausmaß gezogen bzw. umgesetzt werden, wie das sinnvoll wäre. Mit anderen Worten: Die Befunde und Ansatzpunkte sind längst klar, die Diagnose also hinreichend erfolgt, jedoch fehlt es offensichtlich ein wenig an der Therapie. Oder anders und plakativ gesagt: Die Kuh hat man schon mehrfach und regelmäßig gewogen, das Futter ist allerdings noch nicht in der Qualität vorhanden, dass sie bessere und mehr Milch geben könnte.

Natürlich wurden einzelne Maßnahmen durchaus gesetzt, jedoch sind sie nicht so weitreichend, dass dies zu einer nachhaltigen Verbesserung beigetragen hätte. Betrachten wir deshalb die „Haben- und die Sollseite“ ein wenig genauer:

Nach dem ersten „PISA-Schock“ 2004 (zumindest in Österreich und Deutschland) ließ sich nicht mehr leugnen, dass etwa die österreichischen Schülerinnen und Schüler im Durchschnitt durchaus nur mittelmäßige Leserinnen und Leser sind. Eine Reihe von Maßnahmen wurde konsequenterweise getroffen. So wurden seitens des Bildungsministeriums zahlreiche Leseinitiativen zentral und auch in den Bundesländern initiiert. Es gab unter dem Slogan „Nur wer lesen kann, kann auch lernen“ neben der verstärkten Lehreraus- und -fortbildung in diesem Bereich etwa auch Prämien für innovative Leseprojekte an Schulen, ebenso wie die flächendeckende Einführung des Salzburger Lesescreenings in den Volksschulen und auch weiterführenden Schulen, um basale Lesedefizite rechtzeitig erkennen und gegensteuern zu können (BMBWK, 2005). Und die Erfolge stellten und stellen sich auch langsam ein. So verzeichnete die Lesekompetenz bei der PISA-Testung (2012) einen erkennbaren „Aufschwung“ im Vergleich zu PISA 2009 (Bifie, 2013). Weiters wurde damals die sogenannte „vorgezogene Schuleinschreibung“ verankert: Schülerinnen und Schüler, die sich traditionell für die erste Klasse der Volksschule einschrieben, wurden nunmehr nicht erst im März in die Schulen gebeten, sondern bereits im Herbst, um frühzeitig zu erkennen, wo es noch Defizite, beispielsweise in der Unterrichtssprache, gab, und um noch länger Zeit zu haben, die Betroffenen vor dem Schuleintritt zu fördern. Die Kinder wurden mit sogenannten „Sprachtickets“ ausgestattet, einer Art Gutscheine, den sie in Kindergärten oder anderen Institutionen einlösen konnten, um eine gewisse Sprachförderung zu erhalten (Stadt Wien, 2008). Leider war die Organisation dieser Maßnahme relativ überstürzt und derart ineffektiv, dass sie nicht den gewünschten Erfolg zeigte und auch bald wieder eingestellt wurde (Wiener Landtag, 2008, S. 34). Aber immerhin wurde das Problem der z.T. mangelnden Kompetenz in der Unterrichtssprache erkannt und auch angegangen. Ein

Meilenstein in dieser Hinsicht war wohl die bald darauf folgende österreichweite Einführung des verpflichtenden und (zumindest vormittags) gratis angebotenen letzten Kindergartenjahres (Land Niederösterreich, 2014). Das war insofern bemerkenswert, als bis dahin in einigen Bundesländern der Kindergarten kostenpflichtig war und aus diesem Grund eine nicht geringe Zahl an Kindern nicht den Kindergarten besuchte, und dabei gerade jene, die die Förderung in Deutsch am dringendsten gebraucht hätten. Klar ist natürlich in diesem Zusammenhang, dass vermutlich ein Jahr alleine auch nicht ausreicht, um die vorhandenen Defizite auszugleichen, weshalb auch immer wieder Stimmen laut wurden und werden, die den Gratis- und verpflichtenden Kindergarten generell verlangen bzw. fordern, ihn zumindest um ein Jahr auszudehnen (ÖIF, 2014). Und obwohl das einige Bundesländer, wie z.B. Niederösterreich, zumindest für den Vormittagsbereich, bereits seit Längerem umgesetzt haben, ist evident, dass dies natürlich eine finanzielle Frage ist. Hier geht es etwa darum, genug Kindergartenplätze anbieten zu können, was nicht nur aus baulicher und infrastruktureller Sicht nicht einfach ist, sondern auch eine deutlich größere Zahl an Kindergartenpädagoginnen und -pädagogen verlangt, die derzeit nicht in Sicht ist. Das mag auch daran liegen, dass es im Rahmen der „PädagogInnenbildung NEU“ nicht gelungen ist, so wie in anderen Ländern, die Ausbildung dieses Personenkreises auf universitärer Ebene zu verankern, was dem Image der Berufsgruppe nicht sehr zuträglich ist (Plattform Educare, 2014). Aber gerade eine professionelle Ausbildung in diesem Bereich und ein verstärkter Fokus auf die vorschulische Bildung scheinen Schlüsselfaktoren für ein erfolgreiches Bildungssystem zu sein, wie internationale Vergleichsstudien immer wieder zeigen (Bacher, 2008, S. 18; Salcher, 2008, S. 199). Auch die Lehrerbildung generell an den Universitäten anzusiedeln, war und ist ein Ziel der Politik, da es aus internationaler Perspektive immer noch einen Anachronismus darstellt, sich etwa zwei verschiedene Institutionen zu „leisten“, die für die Ausbildung der Pflichtschullehrkräfte und jener Lehrerinnen und Lehrer an höheren Schulen verantwortlich sind, nämlich Pädagogische Hochschulen einerseits und Universitäten andererseits. Hier ist diese gemeinsame Lehrerbildung zwar auf dem Weg (BMBF, 2013), jedoch wirft die Umsetzung derzeit z.T. noch mehr ungeklärte Fragen auf als sie Lösungen anbietet (Universitäre Plattform für LehrerInnenbildung, 2013). Natürlich ist dies ein sehr heikler Bereich, der in die Befindlichkeiten der gewachsenen Institutionen einerseits und der Verantwortungsträger/innen andererseits hinein spielt. Denn weder sind die Zuständigkeiten für beide Bereiche ausreichend geklärt, immerhin sind auch zwei Ministerien dafür zuständig, noch sind derzeit, verfolgt man die mediale Berichterstattung, die politischen Entscheidungsträger/innen (etwa das Bildungsministerium oder die Länder) bereit, Macht abzugeben (was etwa die Personalhoheit betrifft), was natürlich wieder die bereits oben kritisierten Doppelgleisigkeiten und Mehrfachzuständigkeiten „begünstigt“ und letztlich auch für eine – nach wie vor – relativ stark aufgeblähte Verwaltung sorgt. Und gerade das bindet Geldmittel, die an den Schulen so dringend gebraucht werden. Dies wurde auch schon hinlänglich oft kritisiert, da uns hinsichtlich des Bildungssystems „erfolgreiche“ Länder durchaus vorzeigen, wie man es schafft, die aufgewendeten budgetären Mittel direkt den Betroffenen, nämlich den Schülerinnen und Schülern und Lehrkräften, zur Verfügung zu stellen, d.h. in die Klassen zu bringen. Und das ist sicher ein Problembereich, der in Österreich noch längst nicht gelöst ist. So wendet unser Land mehr Geld für Bildung auf als das hoch gelobte Finnland, während es dort allerdings das Schulsystem aufgrund „schlanker“ Strukturen offenbar sehr effektiv schafft, deutlich mehr Unterstützungssysteme an und in die Schulen zu bringen (Salcher, 2008, S. 195). Machen wir das doch einmal an Finnland fest. Dort gibt es an den meisten Schulen:

- Speziallehrer/innen (besonders für leistungsschwache Schüler/innen),
- Schulassistentinnen und -assistenten (zur Unterstützung),
- Sprachlehrer/innen (zur Förderung der Mutter- und Zielsprache),
- fixe Vertretungsreserven (für Supplierungen, damit kein qualifizierter Unterricht ausfällt),
- Schulpsychologinnen und -psychologen,
- Schulärztinnen und -ärzte bzw. Schul(kranken)schwestern,
- Kuratorinnen und Kuratoren (Sozialpädagoginnen und -pädagogen),
- geringe Schulgrößen (40 % aller Schulen haben weniger als 50, nur 3 % mehr als 500 Schüler/innen)
- kleine Klassen (im Schnitt ca. 17 Schüler/innen),
- Küchenpersonal (Jeder finnische Schüler und jede Schülerin erhält jeden Tag ein Mittagessen gratis)

- gratis Arbeitsmaterialien und Schulbücher (von Freyermann, 2002).

Davon kann man hierzulande nur träumen. Dies ist ein Umstand, der nicht nur von der Lehrgewerkschaft immer wieder heftig kritisiert wird. So zeigen Studien deutlich, dass etwa administratives (Unterstützungs-) Personal an Österreichs Schulen im internationalen Vergleich sehr unterdurchschnittlich vorhanden ist. Es fehlt – etwa im Gegensatz zu Finnland – auch an qualifiziertem Personal, etwa was Sozialpädagoginnen und Sozialpädagogen, Psychologinnen und Psychologen oder auch Assistenzlehrkräfte, etwa Sprachlehrerinnen und Sprachlehrer, betrifft (APS, 2014).

Noch gänzlich ungeklärt ist die Frage, wie es unser Schulsystem schaffen könnte, die kontinuierlich produzierte Ungleichheit bzw. signifikant hohe „Vererbung“ von Bildungsabschlüssen zu reduzieren (Bruneforth/Herzog-Punzenberger/Lassnig, 2012, S. 8 und 11-12) Preisendörfer (2008) stellt in seinem Buch „Das Bildungsprivileg“ sogar die These auf, dass Chancengleichheit, entgegen aller Beteuerungen, sogar gesellschaftlich (inoffiziell natürlich) unerwünscht ist. Diese Argumentation anzuziehen, würde natürlich den Rahmen dieser Abhandlung deutlich sprengen. Selbstverständlich hat aber die gesellschaftlich produzierte Ungleichheit auch etwas mit der frühen Selektion zu tun (Bruneforth/Herzog-Punzenberger/Lassnig, 2012, S. 24-25). Österreich ist, neben Deutschland, eines der wenigen Länder, die die Kinder mit zehn Jahren in (zumindest) zwei unterschiedliche Schultypen selektieren. Und gerade das mag ein wesentlicher Mitgrund sein, warum (bereits beim Schuleintritt vorhandene) Defizite und Ungleichheiten nicht (rechtzeitig bzw. gar nicht) ausgeglichen werden können (Bacher, 2008, S. 18; Preisendörfer, 2008). Nun sind Gesamtschulsysteme, wie schon ausgeführt, sowohl bei den PISA-Siegern als auch bei den PISA-Verlierern vertreten, was darauf verweist, dass die Einführung eines gesamtschulischen Systems bzw. einer späteren Selektion nicht unbedingt der (alleinige) Erfolgsfaktor für ein Schulsystem sein dürfte, jedoch zeigen vergleichbare Länder durchaus vor, dass ein späterer Selektionszeitpunkt den Einfluss des familiären Umfelds bzw. auch des Bildungsgrads der Eltern auf den Schulerfolg der Kinder deutlich reduziert (Bacher, 2008, S. 20). Hier wäre also unmittelbarer Handlungsbedarf gegeben, auch da die Bildungswissenschaft schon seit Langem vor dieser frühen Selektion und ihren Nachteilen warnt (z.B. Olechowski/Persy, 1993). Allein die Gesamtschule einzuführen, also quasi lediglich die Türschilder an den Schulen auszutauschen, würde allerdings den gewünschten Erfolg genauso wenig bringen (Die Presse, 2007). Dazu bräuchte es eine Reihe zusätzlicher (innerschulischer) Maßnahmen, die sich vor allem auf die Unterrichtsqualität beziehen; und hier wäre es durchaus lohnenswert, den Blick über den nationalen Tellerrand zu wagen (Figl, 2013).

Der wohl effektivste ungleichheitsreduzierende Systemfaktor, neben der frühen Selektion und der schulischen Frühförderung, dürfte jedenfalls, internationalen Befunden zufolge, die Einführung von echten (!) ganztägigen Schulformen sein, bei denen quasi das Lernen über den Tag verteilt ist und sich mit Übungs- und Erholungsphasen (angelehnt an den Biorhythmus) abwechselt. Erfolgreiche Schulsysteme zeigen vor, dass dies nicht auf Kosten der guten Schülerinnen und Schüler geht (übrigens wie auch bei guten Gesamtschulen nicht) und dass die benachteiligten Kinder stark profitieren (Bacher, 2008, S. 18). D.h. die derzeit produzierte große Heterogenität in den Leistungen der Kinder müsste deutlich entschärft werden. Natürlich ist auch das eine Kostenfrage und bedürfte eines enormen nationalen Schulerschlusses (immerhin sind etwa die Schulerhalter im Pflichtschulbereich die Kommunen, die jetzt schon vielerorts finanziell „am Rande des Abgrunds“ stehen). Allerdings wäre das eine sehr effektive Maßnahme, deren Erfolg sich freilich auch erst langfristig einstellen würde, wie internationale Beispiele zeigen.

Es bleibt also für unser Schulsystem auf der To-do-Liste einiges übrig, es liegt Vieles am Tisch, was man angehen könnte und müsste, jetzt bedarf es nur eines bildungspolitischen Konsenses und eines politischen Willens, diese Maßnahmen zu ergreifen; andernfalls wird es wieder – um ein Bild zu zeichnen – nur oberflächliche „Schönheitsoperationen“ im Schulsystem geben, die den Menschen als solchen jedoch nicht verändern oder verbessern.

2.4 Tests als Trendsetter, Normierungs- und Standardisierungsinstrument – oder: Wie man den Bildungsbegriff erfolgreich amputiert

Wie schon oben ausgeführt, bilden Tests, ob national oder international, Bildung nur in einem sehr eingeschränkten Segment ab. Das, was als Bildung oder Bildungsstandard bezeichnet wird, ist in der Regel das, was die Testautorinnen und -autoren als solches definieren. Und hier besteht der Verdacht, dass der Faktor „Testability“, also die Testbarkeit, eine nicht unbedeutende Rolle spielt: Getestet wird nicht nur das, was

standardmäßig vorgegeben wird, sondern auch das, was sich (relativ) leicht testen lässt. Oft wird in diesem Zusammenhang auch kritisiert, dass standardisierte Tests nur eine Kompetenz zweifelsfrei feststellen (lassen): Ob Schülerinnen und Schüler den Standardtest schaffen, letztlich wird in dieser Hinsicht nur deren Testtauglichkeit getestet (Wagner, 2010, S. 70). Betrachtet man etwa die Aufgabenbeispiele bzw. Testformate und -items der Bildungsstandards, dann orientieren sie sich stark an der Quantifizierbarkeit der Ergebnisse (Bifie, 2014a), was immer wieder auch Kritik seitens der Didaktikerinnen und Didaktiker provoziert (von der Groeben, 2004, S. 689-690; Liessmann, 2006, S. 77; Lusser, 2010, S. 301), z.B. dass „die meisten Kompetenzen beziehungsweise grundlegenden Fertigkeiten des literarischen Verstehens (...) überhaupt nur schwer (formal) zu überprüfen“ (Leubner/Saupe 208, S. 13) sind, wenn man etwa die Standards aus Deutsch betrachtet. Vieles, was in den Schulen unterrichtet wird und werden soll, lässt sich eben schwer bis gar nicht (bzw. nicht ohne großen Aufwand) in Testformaten abbilden. So werden z.B. bei den Bildungsstandards der Volksschule (Bifie, 2014h) nur Deutsch bzw. Mathematik getestet („gemessen“ ist mit Sicherheit eine Anmaßung), und im Bereich Deutsch wird etwa der Bereich Sprechen „aus organisatorischen Gründen“ nicht flächendeckend erhoben (Bifie, 2014f), womit wichtige Informationen über einen bedeutenden Teilbereich stark ausgeklammert werden. Schon aus diesem Grund ist ein Rückschluss auf die Deutschkompetenz (die ja zumindest auch aus Schreiben und Sprechen besteht) der Testees nicht wirklich möglich – und schon gar nicht eine Generalisierung der Ergebnisse auf Bildung bzw. den Erfolg eines Bildungssystems. Die Bildungsstandards auf der achten Schulstufe testen ebenfalls nur die Kompetenzen (bzw. Teilbereiche) in Deutsch, Mathematik und Lebende Fremdsprache (Bifie, 2014h) – wieder einen relativ schmalen – wenn auch zugegebenermaßen bedeutsamen – Ausschnitt aus dem Fächerkanon. Insbesondere musisch-kreative Gegenstände wie Musikerziehung, Werkerziehung oder Bewegung und Sport, um nur einige zu nennen, werden gar nicht in den Blick genommen, ganz zu schweigen von anderen wichtigen Bildungsbereichen, wie etwa Persönlichkeits-, Sozial- oder Methodenkompetenz (Salcher, 2008, S. 202-203). Auch die Werteerziehung oder der Erziehungsauftrag von Schule insgesamt werden in diesem Zusammenhang völlig ausgeblendet. Diese Einschränkung lässt natürlich eine Verallgemeinerung der Ergebnisse bzw. eine Übertragung auf das Bildungswesen (und dessen Leistungsfähigkeit) insgesamt in keinster Weise zu. Allein die Betrachtung des §2 Schulorganisationsgesetz, also des Zielparagrafen des österreichischen Schulsystems (BMBF, 2014), müsste erkennen lassen, dass an unseren Schulen viel mehr unterrichtet bzw. vermittelt werden soll als die reine Fachkompetenz (in ausgewählten Fächern). Dort heißt es nämlich:

„§ 2. Aufgabe der österreichischen Schule

(1) Die österreichische Schule hat die Aufgabe, an der Entwicklung der Anlagen der Jugend nach den *sittlichen, religiösen und sozialen Werten* sowie nach den *Werten des Wahren, Guten und Schönen* durch einen ihrer Entwicklungsstufe und ihrem Bildungsweg entsprechenden Unterricht mitzuwirken. Sie hat die Jugend mit dem für das Leben und den künftigen Beruf erforderlichen Wissen und Können auszustatten und zum *selbsttätigen Bildungserwerb* zu erziehen.

Die jungen Menschen sollen zu *gesunden, arbeitstüchtigen, pflichttreuen und verantwortungsbewußten* Gliedern der Gesellschaft und *Bürgern der demokratischen und bundesstaatlichen Republik Österreich* herangebildet werden. Sie sollen zu *selbständigem Urteil und sozialem Verständnis* geführt, *dem politischen und weltanschaulichen Denken anderer aufgeschlossen* sowie befähigt werden, *am Wirtschafts- und Kulturleben Österreichs, Europas und der Welt Anteil zu nehmen* und in *Freiheits- und Friedensliebe* an den *gemeinsamen Aufgaben der Menschheit mitzuwirken*“ (BMBF, 2014).

Nun raten Sie einmal, welche dieser Ziele und Kernaufgaben von den Standardüberprüfungen NICHT erfasst werden: Richtig – zumindest die kursiv gestellten Bereiche...betrachtet man die Bildungsstandards von dieser Seite, dann sind sie eine glatte Themenverfehlung!

Die Gefahr besteht nun natürlich darin, dass gerade testmäßig abgefragte Gegenstände und Bereiche deshalb in der Praxis „überbelichtet“ und andere, die nicht überprüft werden, „unterbelichtet“ werden (Gruber, 2004, S. 675; Kraus, 2014, S. 15). Die (Schul-) Autonomiebewegung in den 90er Jahren in Österreich hat ja schon gezeigt, dass dann unter Umständen einzelne Fächer quasi „unter die Räder kommen“ und anderen verstärkte Aufmerksamkeit geschenkt wird. Wenn nun Vorgesetzte und/oder Bildungsverantwortliche (ungerechtfertigterweise) noch dazu die Ergebnisse von PISA & Co dafür verwenden, um den Output unseres Schulsystems in Frage zu stellen oder gar die Lehrerleistung daran messen, besteht die große Gefahr, dass gerade Direktorinnen und Direktoren bzw. Lehrkräfte, die ja um ihre Reputation bzw. jene ihrer Schule fürchten, ganz intensiv die testrelevanten Bereiche forcieren und andere „außen vor lassen“. Man will ja schließlich „nicht schlecht dastehen“ und Kritik „von oben“ bekommen. Eine „Ausdünnung“ des Fächer- bzw.

Bildungskanons und eine Verengung auf testrelevante Bereiche könnte ebenso die Folge sein wie das bekannte „Teaching to the test“ (Messner, 2004, S. 702). Dass damit aber einer umfassenden Bildung kein guter Dienst erwiesen wird, scheint klar (Meyer/Zahedi, 2014).

Nun ist seit jeher umstritten bzw. nicht endgültig beantwortet bzw. beantwortbar, was unter Bildung zu verstehen ist bzw. was sie ausmacht; Generationen von Pädagoginnen und Pädagogen bzw. Bildungswissenschaftlerinnen und -schaffern haben sich dieser Frage gewidmet und viele unterschiedliche Antworten dazu geliefert. Sie reichen von sehr detaillierten Beschreibungs- und Definitionsversuchen (Schwendenwein, 1993) über sehr idealistische, humanistische Konzepte (Liessmann, 2006, S. 54) bis hin zum Versuch, Bildung von Produktion zu unterscheiden und das Spannungsfeld zwischen Bildung und ihrer Objektivierbarkeit auszuloten (Messner, 2004, S. 693). Dieser Diskurs kann an dieser Stelle aus Platzgründen natürlich nicht einmal ansatzweise vorgestellt werden. Einig sind sich alle namhaften Expertinnen und Experten allerdings darin, dass etwa das, was PISA & Co, „messen“, NICHT Bildung ist, jedenfalls nicht in einem umfassenden Sinn, bzw. dass zentrale Testungen eben gerade nicht das messen, was sie zu messen vorgeben, nämlich Bildung (Liessmann, 2006, S. 92). Allein das Wort „Bildungsstandards“ ist wohl eine Anmaßung, die suggeriert, dass Bildung gemessen wird; dabei handelt es sich, wie schon gezeigt, maximal um Ausschnitte dessen, was gemeinhin unter Bildung verstanden wird bzw. werden kann. Deshalb wurde dieser Aspekt in der einschlägigen Wissenschaft auch hinlänglich kritisiert. Dennoch hat diese Kritik kaum Auswirkung auf die Interpretation der Testergebnisse seitens der Bildungsverantwortlichen, Politiker/innen oder der Medien. Und dies muss aus Sicht der Pädagogik als sehr bedauerlich bezeichnet werden. Tests werden somit zu Trendsettern, zu Normierungs- und Standardisierungsinstrumenten (Altrichter/Schratz, 2004, S. 642), mit denen Bildung „erfolgreich amputiert“ bzw. reduziert wird (von der Groeben 2004, S. 692). Gerade der normative Anspruch (Liessmann, 2006, S. 86) solcher Tests muss uns hellhörig werden lassen. „Was sich nach den ersten Testreihen unter der Hand abzeichnete, ist nun beim dritten Test schon zum offiziellen Programm geworden: Die Schulen hatten sich besser auf PISA vorzubereiten, die Lehrer sollten die Schüler für die zu erwartenden Aufgaben trainieren, ungeachtet dessen, ob das mit den geltenden Lehrplänen vereinbar ist oder nicht. Die Schulen, wie immer sie organisiert sein mögen und wie immer das Milieu aussieht, in dem sie agieren, werden damit zu Trainingsstätten für die heimlichen Lehrpläne der OECD-Ideologen“ (Liessmann, 2006, S. 86).

Nun ist es nicht so, dass etwa Bildungsstandards im Gehirn eines „schulpolitischen Diktators“ entstehen, der alle Betroffenen gleich schalten möchte, sondern natürlich folgte deren Entwicklung in Österreich – wie oben schon ausgeführt – einem internationalen Trend, im Gefolge dessen Vertreterinnen und Vertreter des Unterrichtsministeriums, der Landes- und Bezirksschulräte, des BIFIE, Didaktikerinnen und Didaktiker bzw. Fachwissenschaftlerinnen und -wissenschaftler von Universitäten und Pädagogischen Hochschulen, Bildungswissenschaftlerinnen und -wissenschaftler, Psychometrikerinnen und Psychometriker, praktisch tätige Lehrkräfte u.v.a.m. in die Entwicklung der österreichischen Bildungsstandards, der Kompetenzmodelle und auch Testinstrumente einbezogen wurden. Es war also durchaus gewollt, eine so einschneidende Maßnahme auf eine möglichst breite Basis zu stellen. Dennoch ist es bezeichnend, dass weder auf der Homepage des Unterrichtsministeriums noch auf jener des BIFIE genaue Angaben dazu gemacht werden, wieso genau jene Standards ausgewählt wurden, die eben ausgewählt wurden, wie es quasi dazu kam, dass aus den Lehrplänen genau jene zu erreichenden Lernergebnisse standardmäßig festgelegt wurden, die eben festgelegt wurden. Dieser Zwischenschritt, also die Begründung für diese Auswahl, fehlt zur Gänze. Natürlich kann man davon ausgehen, dass sich viele schlaue Leute Gedanken darüber gemacht haben, aber expliziert wurden diese Gedanken nicht (wirklich). Gerade das wäre aus bildungswissenschaftlicher Sicht jedoch höchst interessant: Wie kam es zu dieser (konkreten) Normierung? Welche Entscheidungsprozesse standen und stehen dahinter? Welche Ziele waren handlungsleitend? Diese werden z.T. zwar angesprochen, aber auf einer sehr abstrakten und allgemeinen Ebene.

Gleiches gilt auch für die bekannten internationalen Tests, wie PISA, PIRLS oder TIMSS. Zwar wird auf der Homepage der OECD sehr genau ausgeführt, dass Vertreterinnen und Vertreter aus allen Teilnehmerländern und auch internationale Expertinnen und Experten aus allen drei testrelevanten u.a. Domänen an der Konzeption, Administration und Auswertung dieser Tests beteiligt sind, jedoch fehlen transparent gemachte Hinweise darüber, wie die einzelnen Gremien die entsprechenden Fragen und Aufgabenbeispiele auf Basis welcher Überlegungen und Zielsetzungen konkret entwickelten (OECD, 2014). Man gewinnt den Eindruck, es soll vermittelt werden, dass allein die Einbeziehung aller betroffenen Länder und auch von renommierten Expertinnen und Experten dafür garantiert, dass alles seine Richtigkeit hat. Aus (bildungs)wissenschaftlicher Sicht sind diese Informationen allerdings unbefriedigend und unzureichend.

Nun mag es durchaus hinterfragenswert sein, ob diese Evaluationen von Teilbereichen der Bildung, und mehr sind sie nicht, überhaupt der erfahrungswissenschaftlichen Pädagogik zuzurechnen sind. Wenn man diese Testungen allerdings in eine pädagogische Richtung einordnen möchte, dann sind sie zweifelsfrei am ehesten der empirischen Pädagogik bzw. Bildungswissenschaft zuzuschreiben. Tut man das, dann zeigen sich erstaunliche Parallelen zur wissenschaftlichen Pädagogik, wie Brezinka (1971) die Erziehungswissenschaft sieht, bzw. zum „Normenproblem“ in der erfahrungswissenschaftlichen Pädagogik (Ruhloff, 1979). Hier nur exemplarisch drei Beispiele:

Dem PISA-Konsortium bzw. den Bildungsstandards-Verantwortlichen geht es – fast wäre man geneigt zu sagen „im Geiste Brezinkas“ – ebenso um die Beschreibung (bzw. Erhebung) und kausale Erklärung pädagogisch bedeutsamer Tatsachen zur Verbesserung der pädagogischen bzw. – in diesem Fall – der schulischen oder bildungspolitischen Praxis bzw. der Hilfe bei der Formulierung von pädagogischen Aufgaben, im vorliegenden Fall eben z.B. von Aufgaben und Maßnahmen von Bildungssystemen (Ruhloff 1979, S. 73-74; OECD, 2014). Insofern ist sowohl der Anspruch der PISA-Verantwortlichen als auch jener der erfahrungswissenschaftlichen Pädagogik in der Tradition Brezinkas technologischer Art. Man möchte ein Wissen bereitstellen (heute würde man das wohl evidenzbasiert nennen), um einerseits pädagogische oder andererseits bildungspolitische Probleme zu lösen bzw. Entscheidungen zu unterstützen.

Weiters möchte man seitens der Testverantwortlichen Wirkgesetzmäßigkeiten, z.B. hinsichtlich des Zusammenhangs zwischen Migrationshintergrund, Geschlecht, sozioökonomischem Status oder bestimmten Einstellungen und den Testleistungen, aufdecken (OECD, 2014), was auch stets ein Ziel der erfahrungswissenschaftlichen Pädagogik war (Ruhloff, 1979, S. 77).

Ein normatives Problem stellt sich für die Konstrukteurinnen und Konstrukteure der angesprochenen nationalen und internationalen Tests gar nicht (Ruhloff, 1979, S. 72; OECD, 2014). Natürlich ist es nicht so wie bei Brezinka (1971), wo die Erziehungswissenschaft als Erfahrungswissenschaft losgelöst bzw. unterschieden ist bzw. wird von der Erziehungsphilosophie und der praktischen Pädagogik und den Anspruch erhebt, normenfrei zu sein, sondern die Befürworterinnen und Befürworter bzw. Proponentinnen und Proponenten von PISA & Co erheben sogar den Anspruch der Normativität und Standardisierung zum Zwecke der Vergleichbarkeit (OECD, 2014). Insofern handelt es sich hier – im Gegensatz zu Brezinka (1971) – quasi um ein „inverses normatives Problem“, dergestalt, dass die den Testungen unterliegenden Normen, Ziele, Handlungsmotive oder Prinzipien zwar akzeptiert, allerdings nicht ausreichend kritisch hinterfragt oder gar hinreichend begründet werden. In diesem Sinne ist das Normenproblem ebenso (scheinbar) isoliert und minimiert, wie Ruhloff (1979, S. 83) es bezeichnen würde, jedenfalls ist es in den einschlägigen Darstellungen der Institutionen, sei es die OECD oder das BIFIE, explizit ausgeklammert. Und ebenso wie Ruhloff (1979) zu Recht kritisiert, dass der von Brezinka (1971) proponierten Erziehungswissenschaft sehr wohl eine Normativität inhärent ist, dass weiters der Erziehungsbegriff Brezinkas nicht so objektiv ist, wie von ihm erhofft, dass beides aber quasi von den Kritisierten negiert wird, muss der OECD und auch den Verantwortlichen nationaler Testungen unterstellt werden, dass ihr Tun nie frei von Ideologie und Normen ist; man möchte sich zwar den „Anstrich“ der Objektivität und Wissenschaftlichkeit geben, immerhin ist das BIFIE ja z.B. auch ein Institut zur Bildungsforschung, jedoch kann das alleine schon deshalb nicht gelingen, weil dieses Institut eine dem Unterrichtsministerium unterstellte Einrichtung ist, die natürlich die Vorgaben der Politik umzusetzen hat, und das betrifft auch, was z.B. beforscht wird, wie es dargestellt wird und was überhaupt publiziert wird oder werden darf. Ebenso ist die OECD mit Sicherheit keine „neutrale“ Einrichtung im wissenschaftlichen Sinn, sondern hat schon vom Selbstverständnis her wirtschaftliche Interessen. Und wenn schon die akademisch angesiedelte Erziehungswissenschaft nicht normenfrei sein kann, wie soll es dann Institutionen gelingen, die gewissen Intentionen und Interessen ihrer Auftrag- oder Geldgeber/innen dienen?

Es besteht generell der dringende Verdacht, dass die Testeuphorie in direktem Zusammenhang mit einer allgemeinen Ökonomisierung der Gesellschaft und Bildung steht (Salcher, 2008, S. 203-204; Schirlbauer, 2005, S. 201; Preisendörfer, 2008, S. 176; Meyer/Zahedi, 2014). „Von Anfang an stand der Gedanke der Bewertung und Reihung in Verbindung mit dem Paradigma betriebswirtschaftlichen Denkens, das aus Schulen und Universitäten Unternehmen machen wollte, die an ihren marktorientierten Ergebnissen zu messen seien“ (Liessmann 2006, S.81). Und es mag – wie schon gesagt – kein Zufall sein, dass etwa die PISA-Testungen von der OECD verantwortet werden, einem Zusammenschluss von Staaten, die sich zu Marktwirtschaft und wirtschaftlicher Entwicklung bekennen. Nun droht gerade in Zeiten einer allgemeinen Budgetknappheit, wie schon mehrfach angedeutet, dass nur mehr jene Bereiche in den Blick genommen werden, die dem wirtschaftlichen Wachstum dienen, dass nur mehr jene Gegenstände hoch bewertet werden, die marktrelevant sind. Wohl nicht zufällig sprach ein ehemaliger (sehr „schöner“) Finanzminister Österreichs vor gar nicht allzu langer Zeit davon, dass so genannte „Orchideenfächer“ an den Universitäten nicht mehr angeboten werden

sollen bzw. zumindest hinterfragt werden soll, was ihre Wirtschaftlichkeit betrifft (Phillip, 2010). Dies muss einen Aufschrei jedes verantwortungsbewussten Bildungswissenschaftlers und jeder reflektierten Pädagogin provozieren, da ein solcher Gedankengang wohl ein sehr engstirniger und zu kurz gedachter ist. Hätte man etwa die Studienrichtung „Arabistik“ nicht weiter angeboten, wer weiß, wo wir heute hinsichtlich des Wissens um die islamische Welt stünden. Es darf gar keinen Zweifel darüber geben, dass Fächer wie Archäologie, Kunstgeschichte oder auch Bildungswissenschaft, deren ökonomischen Wert man eben gerade nicht direkt messen kann, einen unverzichtbaren Bestandteil unseres Kulturguts darstellen und den Unterschied ausmachen zwischen Ausbildung und Bildung (Schwendenwein, 1993, S. 297). Wer nur für den Arbeitsmarkt ausbildet, wer nur den Wert einer Disziplin an der Ökonomie misst, wird Bürgerinnen und Bürger erhalten, die leicht manipulierbar sind, die des kritischen Denkens nicht mehr mächtig sind und die nur im Dienste der Produktivität stehen – eine äußerst, auch demokratiepolitisch, bedenkliche Sichtweise.

2.5 Von der Utopie der Standardisierung und Messung von Bildung

Wie bereits erwähnt, besteht nicht einmal Konsens darüber, was als Bildung verstanden wird bzw. werden kann. Nun grenzt es aus dieser Perspektive fast an Überheblichkeit, Bildung standardisieren und messen zu wollen. Denn Bildung (in ihrer ganzen Ausprägung und Breite) ist schlicht und ergreifend nicht quantitativ erfassbar und schon gar nicht messbar (Regenbrecht, 2005, S. 17). Damit könnten wir dieses Kapitel auch schon wieder schließen. Es ist aber dennoch wert, ein paar Gedanken lang noch dabei zu verweilen.

Zugegebenermaßen sind es nicht (nur) ein paar Ökonomen, die sich im stillen Kämmerchen um die Konzipierung internationaler und nationaler Testungen Gedanken machen. Bei PISA, ebenso wie bei den Bildungsstandards und allen anderen Testungen, sind es – wie schon erläutert – durchwegs Vertreterinnen und Vertreter aus verschiedensten Ländern und Domänen, die sich in einem Gremium beraten, um möglichst vergleichbare Erhebungen durchzuführen – und dabei sind durchaus Bildungswissenschaftlerinnen und -schafter, Psychologinnen und Psychologen oder auch Expertinnen und Experten aus diversen Fachdidaktiken vertreten (OECD, 2014). Ihnen ist mit Sicherheit gute Absicht und Redlichkeit zu attestieren, davon sollte man bei akademisch und wissenschaftlich ausgebildeten Menschen grundsätzlich ausgehen. Nun lassen sich jedoch Äpfel mit Birnen nur schwer vergleichen (Preisendörfer, 2008, S. 175; Glattauer, 2011, S. 9; Glattauer, 2011, S. 62-75). Die Bildungssysteme der einzelnen Länder sind beispielsweise sehr verschieden. Und wenn nun etwa TIMSS und PIRLS die Kompetenzen auf der vierten Schulstufe miteinander vergleichen, dann muss man wissen, dass es in manchen Ländern eben 8-9jährige sind, die getestet werden, in anderen 9-10jährige. Und gerade in diesem Alter spielt jedes Jahr in der kognitiven Entwicklung eine bedeutsame Rolle (insbesondere auch in Gegenständen wie Mathematik). Der Testvergleich ist also aus dieser Sicht schon nicht fair. Außerdem sind naturgemäß die Stundentafeln und Lehrpläne der einzelnen Länder z.T. sehr unterschiedlich, was etwa die Quantität und die Qualität des Unterrichts betrifft. Auch die Inhalte unterscheiden sich mitunter beträchtlich. Zwar bemühen sich die internationalen Testboards redlich, Grundkompetenzen zu messen, die in allen teilnehmenden Ländern erreicht werden sollten (OECD, 2014), jedoch macht es einen Unterschied, ob man als Schülerin oder Schüler beispielsweise vier Jahre lang nur insgesamt 12, 16 oder gar 20 Stunden in Mathematik geboten bekommt. Und: Schon allein die Geübtheit in externen Testungen macht einen Unterschied. Während es in manchen Ländern, wie Frankreich, Finnland oder mittlerweile auch England, für die Getesteten schon seit Langem üblich und normal ist, dass jemand Externer von außen quasi in den Unterricht kommt und Leistungen misst (Gruber, 2004, S. 675; Ratzki, 2004, S. 681-682), stecken manche Länder, wie Österreich, hier mit Sicherheit noch immer in den Kinderschuhen. Und dieser Effekt darf bei den Ergebnissen und ihren Interpretationen nicht unterschätzt werden. Ganz zu schweigen von den Testformaten: Für manche Länder sind die abgefragten Testitems nichts Ungewöhnliches, für Schülerinnen und Schüler in anderen Ländern ist dies völliges Neuland. Ein Beispiel: Erst mit dem Aufkommen der internationalen Testkultur und auch der Kompetenz- bzw. Outputorientierung (z.B. bei den Bildungsstandards) wandelte sich – zumindest hierzulande – die Sicht auf Unterricht in manchen Gegenständen markant (Messner, 2004, S. 694-697). Die Mathematik soll zunehmend abgehen vom (mechanischen) Rechnen hin zu mathematischem Denken (Fast/Platzgummer, 2010, S. 282; Bifie, 2014b). Doch „unsere“ Lehrerinnen und Lehrer, wie auch Schülerinnen und Schüler, waren darauf – zumindest in den Anfängen – nicht ausreichend vorbereitet. Wenngleich sich dieser Paradigmenwechsel nun auch in der Lehreraus- und -fortbildung langsam durchsetzt, und nebenbei bemerkt auch in den Schulbüchern (als geheimer Lehrplan), wird es wohl noch mindestens ein bis zwei Schüler-Generationen dauern, bis sich diese veränderte Sicht auf Unterricht und seine Ziele in der Praxis durchgesetzt hat. In anderen Ländern wurde dieser Weg schon deutlich früher eingeschlagen. Und das führte dazu, dass „unsere“ Schülerinnen und Schüler

durch die Testaufgaben einfach überfordert waren oder diese schlicht und ergreifend nicht ernst nahmen (Wagner, 2010, S. 44; Glattauer, 2010, S. 13; Glattauer, 2011, S. 62). Dies ist bei der Interpretation der Ergebnisse jedenfalls zu beachten.

Gar nicht vergleichen kann man auch die unterschiedlichen Bildungsaspirationen und den Leistungsdruck in den einzelnen Ländern. Ein paar plakative Beispiele gefällig?

In asiatischen Ländern, wie China oder Japan, ist ein guter Erfolg bei (zentralen) Prüfungen, und somit auch PISA, quasi Ehrensache, ein Versagen eine Schande für die ganze Familie. Die Schülerinnen und Schüler werden auf diese Testungen hin gedrillt, mit Unterricht von zehn Stunden und mehr am Tag – und selbstverständlich auch am Wochenende. So kommen die Heranwachsenden schon gut und gern auf etwa 60 Stunden Unterricht und Nachhilfe pro Woche.

In Südkorea, Taiwan und Shanghai singen die Schülerinnen und Schüler beim Austeilen der Testbögen die Nationalhymne. Ganze Straßenzüge werden für die nationalen Prüfungen abgesperrt, um nur ja keine Beeinträchtigungen der Prüfungssituation zuzulassen.

In China geben Eltern etwa 2000 Euro pro Monat für Schule und Nachhilfeunterricht aus, in Japan gibt es sogar Aufnahmeprüfungen für Nachhilfesschulen (!), wo wieder nur die Besten genommen werden (Glattauer, 2011, S. 62-68 und Trattler, 2014, S. 22).

Dass die Ergebnisse – zumindest in den getesteten Kompetenzen – bei dieser Quantität an Stunden und bei dem Selbstverständnis, mit dem asiatische Schülerinnen und Schüler diese „Prüfungshölle“ auf sich nehmen, durchaus ein wenig besser sind als hierzulande, mag nicht verwundern. Ob der Zweck allerdings den Aufwand heiligt? Nur mal so ein Gedankenexperiment: Wie wäre es denn, wenn man die asiatischen Schülerinnen und Schüler etwa in Bereichen wie Selbst- oder Sozialkompetenz testet. Wären sie da auch PISA-Spitzenländer? Vermutlich eher weniger, wenn man weiß, wie Unterricht in den dortigen Schulen abläuft: Klassen mit bis zu 50 Schülerinnen und Schülern; Frontalunterricht; Pauken; Drill; Disziplin; Sprechen nur, wenn man gefragt wird; macht man – zeigt man auf und wird drangenommen – einen Fehler, wird man bestraft, nicht etwa gelobt, dass man „brav“ mitgearbeitet hat usw. Ja, so kann Unterricht auch funktionieren. Für die Kinder Asiens (und auch die Erwachsenen) ist das völlig selbstverständlich und wir dürfen uns hier auch keine selbstherrliche Kritik dazu anmaßen. Was wir aber dürfen: Eine gewisse Verhältnismäßigkeit bei der Interpretation der PISA-Daten seitens unserer Politik einfordern.

Und: Wenn wir schon Länder miteinander vergleichen, dann müsste man wohl auch andere Aspekte eines erfolgreichen Schulsystems, einer Wirtschaft oder eines ganzen Landes im Blick haben: Wer vergleicht etwa die Arbeitslosen- und Selbstmordrate, den Alkoholkonsum, die Zufriedenheit (z.B. der Schülerinnen und Schüler), die Wirtschaftsleistung, den sozialen Frieden u.dgl. miteinander bzw. bezieht sie in die Bewertung von Bildungssystemen und deren „Output“ ein? (Wagner, 2010, S. 46) Dann lägen manche PISA-Sieger wohl längst nicht mehr vor uns.

Darüber hinaus stellten namhafte Expertinnen und Experten auch eine Reihe an Schwachstellen am Design der Studien, am Sampling und in der Testdurchführung selbst fest. Diese lassen sich wie folgt zusammenfassen:

- grundsätzlich mangelnde Vergleichbarkeit von Ländern (z.B. aufgrund der Quote von Migrantenkindern in den Klassen, unterschiedliche Bildungsaspirationen, verschiedene Prüfungskulturen)
- mangelnde Vergleichbarkeit von Zielsetzungen (von Schule), Inhalten (von Lehrplänen) und Stundentafeln (insbesondere, was die Quantität einzelner Gegenstände betrifft)
- unterschiedliche Anzahl an Schuljahren (und Lebensalter!) zum Testzeitpunkt (macht gerade bei den jüngeren Schülerinnen und Schülern aus entwicklungspsychologischer Sicht sehr viel aus!)
- mangelnde Erfahrung mit zentralen Testungen (externen Prüfungssituationen) generell und (quantitativen) Testformaten und -items speziell in einzelnen Ländern
- Mängel bei der Repräsentativität und Vergleichbarkeit der Samples (Stichprobe), z.B. aufgrund der Ausscheidung bestimmter Gruppen von Schülerinnen und Schülern
- Herkunft und Formulierung der Testitems problematisch
- Unterschiede im Schwierigkeitsgrad der Aufgaben

- unterschiedliche Länge der Texte und Fehler bei Übersetzungen
- Inhomogenität der Fragen
- uvm. (Hopmann/Brinek/Retzl, 2007; Preisendörfer, 2008, S. 176; Glattauer, 2011; Meyer/Zahedi, 2014)

Außerdem wurde generell der Ranglistenfetischismus bzw. -wahn in einzelnen Ländern, der – quasi – Wettbewerbscharakter der Testungen und die bildungspolitische Überbewertung der Ergebnisse hinlänglich kritisiert (Liessmann, 2006, S. 74-103; Preisendörfer, 2008, S. 175). Insgesamt fordern deshalb viele Bildungswissenschaftlerinnen und -wissenschaftler einen gelassenen Umgang mit den Daten. Diese Einsicht wäre den Bildungspolitikern und -politikern vieler Länder, insbesondere auch in Österreich, nur zu wünschen.

3 Zusammenfassung und Ausblick

Der vorliegende Beitrag widmete sich nationalen und internationalen Testungen und der Unmöglichkeit der Standardisierung und Messung von Bildung. Die zu Beginn aufgestellten Forschungsfragen können nunmehr wie folgt in aller gebotenen Kürze – und um Redundanz zu vermeiden – beantwortet werden:

Ad 1) Welche wichtigen nationalen und internationalen Testungen gibt es und was sind ihre Zielsetzungen?

Die wichtigsten und bekanntesten internationalen Testungen, um das (wie wir mittlerweile wissen) unangebrachte Wort „Bildungs“-Studien zu vermeiden, sind TIMSS, PIRLS und PISA und ihr gemeinsames und vorrangiges Ziel ist es, die Leistungen von Schülerinnen und Schülern aus den Teilnehmerländern in ausgewählten Domänen und auf unterschiedlichen Schulstufen miteinander zu vergleichen, um etwa den einzelnen Bildungssystemen Rückmeldungen über die jeweilige Performance zu geben. Gleiches gilt an sich für die nationalen Bildungsstandards, die im Sinne eines System-Monitorings die Leistungen der Schülerinnen und Schüler innerhalb Österreichs überprüfen und den Anspruch erheben, für mehr Bildungsgerechtigkeit und Vergleichbarkeit zu sorgen, zumindest was die teilzentrale Reifeprüfung betrifft.

Ad 2) Was können diese Testungen leisten und wo liegen ihre Grenzen und auch mögliche Gefahren?

Diese Testungen haben eine Reihe von Defiziten des österreichischen Schulsystems zu Tage gefördert, etwa, dass wir in bestimmten Kompetenzbereichen international – trotz hohen finanziellen Inputs – bestenfalls durchschnittlich sind und vergleichsweise wenig effizient, dass wir einen relativ großen Unterschied zwischen den sehr guten und besonders schlechten Schülerinnen und Schülern haben, dass insbesondere die Kinder mit Migrationshintergrund nicht jene Leistungen bringen, die sie bringen könnten und dass es in unserem Bildungssystem nach wie vor zu einer so genannten „Bildungsvererbung“ kommt. Die Grenzen dieser Testungen liegen besonders dort, wo sie den Anspruch erheben, Bildung oder den Erfolg eines Schulsystems insgesamt zu messen, denn diesen lösen sie nicht ein, da nur ein sehr schmaler Ausschnitt dessen getestet wird, was allgemein als Bildung bezeichnet wird bzw. zumindest dessen, was an unseren Schulen in den einzelnen Fächern unterrichtet wird und auch werden soll. Die große Gefahr besteht deshalb u.a. in einer (bewussten oder unbewussten) Fehlinterpretation der Daten und in entsprechenden (unangemessenen oder überzogenen) bildungspolitischen Entscheidungen.

Ad 3) Welche Konsequenzen könnten allenfalls aus den Studien und ihren Ergebnissen auf nationaler Ebene abgeleitet werden?

Abgesehen von den bereits durchgeführten Maßnahmen wie großangelegten Leseinitiativen, dem verpflichtenden letzten Kindergartenjahr oder der vorgezogenen Schuleinschreibung, müssen – vergleicht man Österreich mit „erfolgreichen“ Schulsystemen – noch einige Vorhaben auf der Agenda bleiben und dringend realisiert werden: So wäre es etwa unbedingt erforderlich, eine Verwaltungsreform im Bildungsbereich tatsächlich zu verwirklichen, um jene Mittel und Ressourcen frei zu bekommen, die man besonders in den Schulen dringend benötigt, um etwa Fördermaßnahmen effektiv durchführen zu können und um generell die – im internationalen Vergleich – fehlenden Unterstützungssysteme zu etablieren. Weiters wäre zu überlegen, wie man die vorschulische Förderung noch besser umsetzen kann, sei es durch eine Ausbildung der Kleinkindpädagoginnen und -pädagogen auf hochschulischem Niveau oder sei es durch eine Ausdehnung der Kindergartenpflicht um ein oder zwei weitere Jahre, um nur zwei Beispiele zu nennen. Auch die Nachteile der

frühen schulischen Selektion müsste man in den Griff bekommen, um die Chancenungleichheiten des österreichischen Schulsystems zu reduzieren. Allein durch die Einführung einer Gesamtschule würde man das wohl nicht erreichen, viel eher wären – neben anderen Maßnahmen – (echte) ganztägige Schulformen effektiv, wie Vergleichsstudien immer wieder nahelegen.

Ad 4) Inwiefern fungieren diese Tests als Trendsetter und tragen zu einer Normierung und Einengung von Bildung und Unterricht bei?

Das größte Problem in diesem Zusammenhang ist das bekannte Phänomen „Teaching to the test“; dass also besonders das im Unterricht favorisiert wird, was getestet wird, und dass jene Gegenstände oder Lernbereiche mehr oder weniger „ausgeblendet“ oder zumindest vernachlässigt werden, die nicht testrelevant sind. Es kann also dazu kommen, dass der Unterricht an den Schulen besonders jene Kompetenzen in den Blick nimmt, die überprüft werden, was in weiterer Folge einerseits zu einem ziemlich „schmalspurigen“ Verständnis von Bildung führen könnte, und andererseits dazu, dass viele wertvolle Domänen des Lehrplans einfach „links liegen gelassen“ werden.

Ad 5) Inwiefern ist es überhaupt grundsätzlich möglich, Bildung zu standardisieren und zu (ver)messen? Da schon allein der Bildungsbegriff nicht (leicht) „dingfest“ zu machen ist, scheint es rein logisch ein fast aussichtsloses Unterfangen, Bildung standardisieren oder gar messen zu wollen. Und selbst das, was das PISA-Konsortium oder die Bildungsstandards-Verantwortlichen als besonders wichtige Kompetenzen identifiziert haben, ist letztlich nur sehr bedingt messbar; die Erfahrungen haben sogar gezeigt, dass es schon bei der Formulierung der Standards und Testitems gravierende Probleme gibt, die sich dann bis in die Testdurchführung und zur Interpretation hin durchziehen. Und erst recht problematisch sind oft die (bildungspolitisch) vorgenommenen Ableitungen und geforderten bzw. gezogenen Konsequenzen aus den Testergebnissen. Alles in allem eignen sich nämlich alle groß angelegten Testungen, seien sie national oder international, wie schon oben angesprochen, NICHT für die Evaluation oder Bewertung von Schulsystemen und schon gar nicht von Bildung. Viel zu viele Faktoren determinieren den Erfolg eines Bildungswesens und allein darüber, was als „Erfolg“ bezeichnet wird, besteht längst kein Common Sense.

Für Leserinnen und Leser sollte durch die Lektüre dieser Arbeit einiges klarer sein: Es sollte z.B. klarer sein, wie problematisch besonders die von Nicht-Wissenschaftlern (seien es Politikerinnen und Politiker, Journalistinnen und Journalisten, Schulaufsichtsbeamtinnen und -beamte, Direktorinnen und Direktoren, manche selbsternannte Bildungsexpertinnen und -experten usw.) vorgenommene Hochstilisierung von nationalen und internationalen Testungen und deren Bedeutung bzw. Ergebnissen ist.

Lehrerinnen und Lehrern bleibt zu wünschen, dass ihre direkten Vorgesetzten, etwa mit den Ergebnissen von Bildungsstandardtests in ihren Klassen, ebenso vorsichtig und behutsam umgehen, wie es eigentlich angezeigt wäre.

Und weiters bleibt zu hoffen, dass sich die Bildungspolitik endlich jene Zeit nimmt, die notwendig wäre, um die richtigen Schlüsse aus den Ergebnissen zu ziehen, etwa indem sich manche Verantwortliche das Buch „PISA zufolge PISA“ von Stefan Hopmann & Co sinnerfassend durchlesen.

Literatur

- Aigner, L. (2014). Bildungsstandards: Das zerstört den Unterricht. Online verfügbar unter <http://derstandard.at/1389858751677/Bildungsstandards-Das-zerstoert-den-Unterricht>, geprüft am 07.02.2018.
- Altrichter, H./Schatz, M. (2004). Bildungsstandards und die Weiterentwicklung von Schule und Unterricht. *Erziehung & Unterricht*, 7-8/2004, 630-645.
- APS (Hrsg.). (2014). Hält Schmied Daten unter Verschluss? Online verfügbar unter <http://www.pflichtschullehrer.at/themen/h%C3%A4lt-schmied-daten-unter-verschluss>, geprüft am 10.06.2014.
- Bacher, J. (2008). Ungleichheit im österreichischen Bildungswesen. Aktuelle empirische Befunde. Online verfügbar unter <http://www-oedt.kfunigraz.ac.at/Babylon2/01-Artikel-1/004-Bildungsungleichheit%20in%20Oesterreich-2.pdf>, geprüft am 05.06.2014.
- Barber, M./Mourshed, M. (2007). How the world's best performing school systems come out on top. New York: McKinsey & Company.
- Bear, R. (2005). Bildungsstandards und Leistungsbeurteilung. *Erziehung & Unterricht*, 5-6/2005, 478-484.

- Bifie (Hrsg.). (2013). PISA 2012 – Erste Ergebnisse. Online verfügbar unter https://www.bifie.at/system/files/dl/pisa12_zusammenfassung_ergebnisse_2013-12-03.pdf, geprüft am 07.02.2018.
- Bifie (Hrsg.). (2014a). Häufig gestellte Fragen. Online verfügbar unter <https://www.bifie.at/node/96>, geprüft am 07.02.2018.
- Bifie (Hrsg.). (2014b). Kompetenzen und Modelle. Online verfügbar unter <https://www.bifie.at/node/49>, geprüft am 07.02.2018.
- Bifie (Hrsg.). (2014c). PIRLS. Lesestudie für die 4. Schulstufe. Online verfügbar unter <https://www.bifie.at/pirls>, geprüft am 07.02.2018.
- Bifie (Hrsg.). (2014d). PISA 2003. Online verfügbar unter <https://www.bifie.at/node/93>, geprüft am 05.06.2014.
- Bifie (Hrsg.). (2014e). PISA-Studie für 15-/16-Jährige. Online verfügbar unter <https://www.bifie.at/pisa>, geprüft am 05.06.2014.
- Bifie (Hrsg.). (2014f). Standardüberprüfung Deutsch 4. Schulstufe. Online verfügbar unter <https://www.bifie.at/node/2353>, geprüft am 12.06.2014.
- Bifie (Hrsg.). (2014g). TIMSS: Mathematik und Naturwissenschaft für die 4. Schulstufe. Online verfügbar unter <https://www.bifie.at/timss>, geprüft am 07.02.2018.
- Bifie (Hrsg.) (2014h): Überprüfungskonzept. Online verfügbar unter <https://www.bifie.at/node/57>, geprüft am 07.02.2018.
- BMBF (Hrsg.) (2013): PädagogInnenbildung NEU. Online verfügbar unter <http://www.bmukk.gv.at/lbneu>, geprüft am 10.06.2014.
- BMBF (Hrsg.). (2014). Schulorganisationsgesetz. Online verfügbar unter <http://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009265>, geprüft am 12.06.2014.
- BMBWK (Hrsg.). (2005). Initiative Lesen fördern. Protokoll der Steuerungsgruppe Lesen der Besprechung vom 02.12.2005. Wien: BMBWK.
- BMUK (Hrsg.). (1993). Schulautonomie konkret. Wien: BMUK.
- Braun, H./Weidinger, W. (Hrsg.). (2012). Die standardisierte, kompetenzorientierte Reifeprüfung an AHS: Einblicke und Ausblicke. *Erziehung & Unterricht, 1-2/2012*, .5-83.
- Brezinka, W. (1971). Von der Pädagogik zur Erziehungswissenschaft. Weinheim/Berlin/Basel: Beltz.
- Bruneforth, M. u.a. (Hrsg.). (2016). Nationaler Bildungsbericht Österreich 2015, Band 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen. Graz: Leykam.
- Bruneforth, M./Herzog-Punzenberger, B./Lassnig, L. (Hrsg.). (2012). Nationaler Bildungsbericht Österreich. Indikatoren und Themen im Überblick. Graz: Leykam.
- Die Presse (Hrsg.). (2007). Gesamtschule: Pädagoge warnt vor Schulversuchen. Online verfügbar unter http://diepresse.com/home/politik/innenpolitik/302777/Gesamtschule_Paedagoge-warnt-vor-Schulversuchen, geprüft am 10.06.2014.
- Dobart, A. (2004). Bildungsstandards im Kontext von Qualitätsentwicklung und internationaler Erfahrungen. *Erziehung & Unterricht, 7-8/2004*, 560-565.
- Fast, M./Platzgummer, F. (2010). Bildungsstandards Mathematik 4. *Erziehung & Unterricht 3-4/2010*, 281-290.
- Feller, W. (2015). Die Phrase von der vererbten Bildung. Online verfügbar unter <http://diepresse.com/home/meinung/gastkommentar/4889051/Die-Phrase-von-der-vererbten-Bildung>, geprüft am 07.02.2018.
- Figl, B. (2013). Pisa – ein Test in Schiefelage. Online verfügbar unter <http://bettinafigl.net/pisa-ein-test-in-schieflage/>, geprüft am 07.02.2018.
- Forum Politische Bildung (Hrsg.). (2014). Längsschnitt der österreichischen Schulgeschichte. Online verfügbar unter http://www.politischebildung.com/pdfs/ecker_aw.pdf, geprüft am 07.02.2018.
- Friedl-Lucyshyn, G./Benischek, I. (2010). Implementierungskonzepte konkret dargestellt. *Erziehung & Unterricht, 3-4/2010*, 228-235.
- Gillard, D. (2011). Education in England: A brief history. Online verfügbar unter <http://www.educationengland.org.uk/history/>, geprüft am 05.06.2014.

- Glattauer, N. (2011). Die Pisa-Lüge. Wie unsere Schule wirklich besser wird. Wien: Ueberreuter.
- Gruber, K. H. (2004). Bildungsstandards: „World class“, PISA-Durchschnitt und österreichische Mindeststandards. *Erziehung & Unterricht*, 7-8/2004, 666-677.
- Hameyer, U./Heckt, H. (2004). Nichts als Standards. *Grundschule*, 10/2004, 8-10.
- Hartweger, K. (2016). Experte rät: „Ein eigenes Schulmodell entwerfen.“ Online verfügbar unter http://www.kleinezeitung.at/kaernten/osttirol/peak_osttirol/4997245/Lienz_Experte-raet_Ein-eigenes-Schulmodell-entwerfen?from=suche.intern.portal, geprüft am 07.02.2018.
- Helmke, A./Schrader, F.-W. (2001). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handbuch Pädagogischer Psychologie* (S. 81-91). Weinheim: Beltz.
- Hopmann, S. T./Brinek, G./Retzl, M. (Hrsg.).(2007). PISA zufolge PISA – PISA According to PISA. Wien: Lit.
- IHS (Hrsg.). (2007). Ökonomische Bewertung der Struktur und Effizienz des österreichischen Bildungswesens und seiner Verwaltung. Wien: Institut für Höhere Studien.
- Ingenkamp, K. (1971). Die Fragwürdigkeit der Zensur Umgebung. Weinheim: Beltz.
- Kern, A. (2004). Auf dem Weg zu Bildungsstandards – kurzes Resümee einer mühevollen Spurensuche. *Erziehung & Unterricht*, 7-8/2004, 566-570.
- Kraus, J. (2014). „Ich halte nichts von dieser Testerei.“ *ÖPU-Wochenspiegel*, 13/2014, 15-16.
- Land Niederösterreich (Hrsg.). (2014). Verpflichtendes Kindergartenjahr. Online verfügbar unter http://www.noel.gv.at/Gesellschaft-Soziales/Kinderbetreuung/Kindergaerten/Hausliche_Erziehung.html, geprüft am 09.06.2014.
- Lentner, M. (2008). Die Hauptschule und die Schulautonomie. Gelungene Reaktion auf die Wissensgesellschaft oder weitere Ausdifferenzierung des Bildungswesens. (Diplomarbeit). Johannes Kepler Universität, Linz.
- Leubner, M./Saupe, A. (2008). Textverstehen im Literaturunterricht und Aufgaben. Hohengehren: Schneider.
- Liessmann, K. P. (2006). Theorie der Unbildung. Wien: Paul Zsolnay.
- Lusser, G. (2010). Lesekompetenzforschung und Aufgabenkonstruktion auf der Grundlage von Bildungsstandards – dargestellt für das literarische Lesen in der Grundschule. *Erziehung & Unterricht* 34/2010, 300-310.
- Mayrhofer, E./Schallenberg, E. (2012). Die neue Reifeprüfung an den AHS – eine Chance für die Unterrichtsentwicklung. *Erziehung & Unterricht*, 1-2/2012, 44-60.
- Messner, R. (2004). Was Bildung von Produktion unterscheidet – oder: Die Spannung von Freiheit und Objektivierung und das Projekt der Bildungsstandards. *Erziehung & Unterricht*, 7-8/2004, 693-716.
- Meyer, H.-D./Zahedi, K. (2014). Offener Brief an Andreas Schleicher, OECD, Paris. Online verfügbar unter <http://bildung-wissen.eu/wp-content/uploads/2014/05/offener-brief-schleicher-autorisierte-fassung.pdf>, geprüft am 07.02.2018.
- Mohr, E. (2012). Finnland: Hohe Jugendarbeitslosigkeit trotz PISA-Sieg. Online verfügbar unter <https://www.kommunalnet.at/news/artikel/article/finland-hohe-jugendarbeitslosigkeit-trotz-pisasieg.html?cHash=4dcbdb6007869ce407859cc867e681fd>, geprüft am 07.02.2018.
- Nimmervoll, Lisa (2014). Nach Pannen bei Zentralmatura: Bifie-Direktoren treten zurück. Online verfügbar unter <http://derstandard.at/1399507529022/Das-Staatsinstitut-Bifie-zwischen-Gott-und-Kaiser>, geprüft am 07.02.2018.
- OECD (2014). PISA FAQ. Online verfügbar unter <http://www.oecd.org/pisa/aboutpisa/pisafaq.htm>, geprüft am 12.06.2014.
- ÖIF (Hrsg.). (2014). Staatssekretär Kurz: Zweites verpflichtendes Kindergartenjahr bei Sprachdefiziten. Online verfügbar unter http://www.integrationsfonds.at/nap/staatssekretar_kurz_zweites_verpflichtendes_kindergartenjahr/, geprüft am 10.06.2014.
- Olechowski, R./Persy, E. (1993). Frühe schulische Auslese. Frankfurt am Main: Lang.
- Phillip, N. (2010). Ausgefallene Studienfächer: Kein Artenschutz. Online verfügbar unter http://diepresse.com/home/bildung/universitaet/541866/Ausgefallene-Studienfaecher_Kein-Artenschutz, geprüft am 07.02.2018.

- Plattform EduCare (Hrsg.). (2014). Regierungsklausur: PädagogInnenbildung NEU wieder ohne KindergartenpädagogInnen. Online verfügbar unter <http://www.plattform-educare.org/Presse%2009.11.2012.htm>, geprüft am 10.06.2014.
- Preisendörfer, B. (2008). Das Bildungsprivileg. Warum Chancengleichheit unerwünscht ist. Frankfurt am Main: Eichbor.
- Ratzki, A. (2004). Mehr Leistung durch Standards – aber wer ist dafür verantwortlich? *Erziehung & Unterricht*, 78/2004, 678-683.
- Regenbrecht, A. (2005). Sichern die Bildungsstandards die Bildungsaufgabe der Schule? *PÄD Forum: unterrichten, erziehen*, 2005, 16-22.
- Riegler, G. (2014). Keine Träne. Online verfügbar unter <http://quinecke.wordpress.com/tag/standards/>, geprüft am 05.06.2014.
- Rieß, C./Meließnig, C. (2010). Bildungsstandards oder Vom Haben zum Sein. *Erziehung & Unterricht*, 3-4/2010, 355-361.
- Roth, J. (2014). Finnische Skizzen: Was ist faul im Pisa-Musterland? Online verfügbar unter <http://www.nzz.ch/wissenschaft/bildung/was-ist-faul-im-staate-musterland-1.18254549>, geprüft am 30.06.2014.
- Ruhloff, J. (1979). Das ungelöste Normenproblem der Pädagogik. Heidelberg: Quelle & Meyer.
- Salcher, A. (2008). Der talentierte Schüler und seine Feinde. Salzburg: Ecowin.
- Schatzl, A. (2012). Die standardisierte kompetenzorientierte Reifeprüfung – ein Paradigmenwechsel. *Erziehung & Unterricht*, 1-2/2012, 7-12.
- Schirlbauer, A. (2005). Die Moralpredigt. Destruktive Beiträge zur Pädagogik und Bildungspolitik. Wien: Sonderzahl.
- Schmid, G. (2014). Bildungspolitik nach dem Pinocchio-Prinzip. *ÖPU-Wochenspiegel*, 19/2014, 3.
- Schwendenwein, W. (1993). Theorie des Unterrichtens und Prüfens. Wien: WUV.
- Specht, W./Freudenthaler H. H. (2004). Bildungsstandards – Bedingungen ihrer Wirksamkeit. *Erziehung & Unterricht*, 7-8/2004, 618-629.
- Stadt Wien (Hrsg.). (2008). Vorgezogene Schuleinschreibung in Wien gestartet. Online verfügbar unter <http://www.vienna.at/vorgezogene-schuleinschreibung-in-wien-gestartet/news-20080428-11182890>, geprüft am 09.06.2014.
- Trattler, G. (2014). Examination hell – Nachhilfe am Beispiel Japans. *ÖPU-Nachrichten*, 2/2014, 22.
- Universitäre Plattform für LehrerInnenbildung (Hrsg.). (2013). Stellungnahme. Online verfügbar unter <http://www.upl.or.at/>, geprüft am 10.06.2014.
- Vierlinger, R. (1993). Die offene Schule und ihre Feinde. Beiträge zur Schulentwicklung. Wien: Jugend & Volk.
- von der Groeben, A. (2004). Vorwärts in die Vergangenheit? Oder: Schulqualität gedeiht nicht im Vergleichs- und Kontrolldruck. *Erziehung & Unterricht*, 7-8/2004, 684-692.
- von Freymann, T. (2002). Zur Binnenstruktur des finnischen Schulsystems. Online verfügbar unter <http://www.finland.de/dfgnrw/dfg043a-pisa07.htm>, geprüft am 07.02.2018.
- Wagner, G. (2010). Tatort Pädagogik. Alles Reform – oder was? Wien/Klosterneuburg: Edition Vabene.
- Wiener Landtag (Hrsg.). (2008). 18. Wahlperiode, 17. Sitzung vom 05. Juni 2008. Wörtliches Protokoll. Wien: Stadt Wien.
- Wößmann, L. (2008). Mehrgliedrigkeit des Schulsystems und Chancengleichheit im internationalen Vergleich. *Erziehung & Unterricht*, 7-8/2008, 509-517.

¹ Der vorliegende ist nach den derzeit gültigen Regeln der deutschen Rechtschreibung verfasst. Direkte Zitate aus älteren Quellen werden in der Originalschreibweise übernommen.