

Tanja Müller

Seit nun fast 4 Jahren darf ich an der Seite von Frau Dr. Marianne Häuptle-Barceló an der Professur für die Didaktik der englischen Sprache und Literatur arbeiten und von ihr lernen. Frau Dr. Häuptle-Barceló war mir in all der Zeit eine große Unterstützung, fachlich wie menschlich, und konnte mir Kraft und Zuversicht geben, meinen eigenen Platz im breiten Feld der Wissenschaft zu finden. Sie hat mir vermittelt, dass nicht jeder alles gleich können kann und manches auch gar nicht notwendig ist und man irgendwann an dem Punkt ist, wo man weiß, was man will. Ich danke dir, Marianne, für die Zeit, in der du immer ein aufmunterndes Lächeln für mich übrig hattest. Der folgende Artikel soll dir zeigen, wo ich für mich mein Plätzchen gefunden habe.

Von Gerechtigkeit und Objektivität

Was von Gesellschaftswissenschaftlern schon lange kritisiert worden war, wurde durch PISA bestätigt: die sogenannte Bildungsexpansion bringt lange keine gleichen Bildungschancen. Die Ergebnisse aller, bislang durchgeführten Studien zeigen deutliche Leistungsunterschiede zwischen deutschen Kindern und Kindern mit Migrationshintergrund. Auch der sozioökonomische Status der Eltern hat einen Einfluss auf die Schulleistungen der Schüler. Kinder aus sozial schlechter gestellten Verhältnissen können allein aufgrund der Bildungsexpansion noch keinen sozialen Aufstieg erreichen.

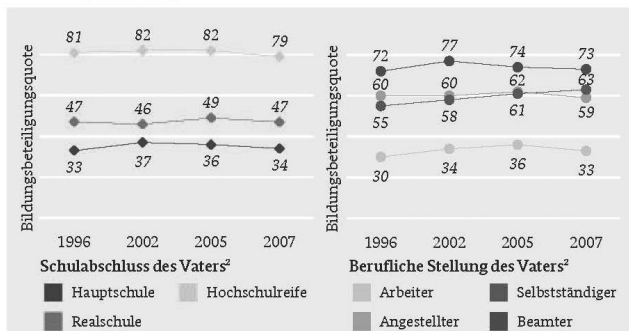
„Tatsächlich ist die Expansion der höheren Bildungseinrichtungen darauf zurückzuführen, dass vor allem die Söhne und Töchter der oberen Bildungs- und Besitzmilieus (...) ihre Beteiligung an der Gymnasial- und Hochschulbildung mehr als verdoppelt haben“ (Vester 2004). Für die unteren Milieus ist dies leider nicht der Fall, sie sind weiterhin stark unterrepräsentiert, wie die folgenden Daten belegen. „Ein zentraler Befund der internationalen Schulleistungsstudien lautet, dass die deutsche Schule so wenig wie kein andere dazu beiträgt, Schülerinnen und Schüler mit schlechten Startchancen zu guten Leistungen zu verhelfen: In keinem anderen Land ist der Schulerfolg der Kinder in dem Maße abhängig von ihrer sozialen Herkunft wie im reichen Deutschland“ (Böttcher 2005, S.7). Die Bundesrepublik Deutschland hat ein föderalistisch aufgebautes dreigliedriges Schulsystem und ein Recht auf einen Kindergartenplatz.

Durch fünf verschiedene Selektionsstufen⁷⁶ verringert sich die Chance von Kindern aus niedrig gebildeten Familien auf einen höheren Bildungsabschluss stark. Die erste Selektionsstufe passiert beim Übertritt von der Grundschule auf eine weiterführende Schule. Zahlen des statistischen Bundesamtes belegen, dass auf den Gymnasien die Arbeiterschichten deutlich unterrepräsentiert sind, obwohl sie den größten Anteil an der Gesamtbevölkerung bilden. Von je 100 15 - 18 Jährigen aus Familien folgender Berufsgruppen besuchen das Gymnasium: 27 Beamtenkinder, 22 Kinder von Selbstständigen, 22 Angestelltenkinder aber nur 11 Kinder aus Arbeiterhaushalten (Statistisches Bundesamt 2003). Ein ähnliches Bild zeigt sich, wenn man nicht die Art des Berufes sondern des Schulabschlusses des Vaters als Vergleichsgröße heranzieht. „Jugendliche ausländischer Staatsangehörigkeit sind den Gymnasien deutlich unterrepräsentiert. Sie bleiben überdurchschnittlich häufig ganz ohne Schulabschluss“ (Bundesministerium für Bildung und Forschung, S.72). Hinsichtlich des Geschlechts ist anzuführen, dass Frauen zwar öfter als Männer die Hochschulzugangsberechtigung erreichen (2004: 39% Männer zu 45% Frauen), aber deutlich seltener eine Hochschule besuchen (vgl. Bundesministerium für Bildung und Forschung, S.91). Dieser Verlauf zeichnet sich während der gesamten Hochschulkarriere weiter ab, Frauen machen nur einen minimalen Anteil bei den W3-Professuren aus. Aber nicht nur das Geschlecht spielt bei der Entscheidung für oder gegen einen Studienbeginn eine Rolle. Wieder sind die Merkmale Herkunftsfamilie und Migrationshintergrund signifikant. „Studienberechtigte aus Familien, in denen mindestens ein Elternteil ein Hochschulstudium abgeschlossen hat, verzichten 2005 lediglich zu 22% auf eine Studienaufnahme, während solche, deren Eltern nicht studiert haben, zu 41% auf ein Studium verzichten“ (Ebenda, S.95). Seit 2002 sind die Zahlen für Kinder ohne studierte Eltern sogar rückläufig, während die Zahlen für die Akademikerkinder weitgehend konstant bleiben. Im Vergleich Hochschulabschluss des Vaters zu Hauptschulabschluss des Vaters wird der Unterschied radikal: Während 2005 71% aller jungen Erwachsenen zwischen 18 und 25 Jahren mit Hochschulzugangsberechtigung und Akademikervater ein Studium aufnahmen, trifft das lediglich auf 19% derer zu, deren Vater

⁷⁶ 1. Schwelle liegt am Übergang von der Grundschule auf eine weiterführende Schule (Realschule, Gymnasium etc.). 2. Schwelle bezeichnet die Übertrittsmöglichkeit nach der 10. Klasse. Die 3. Stufe steht für den Erwerb der Hochschulzugangsberechtigung. 4. Schwelle ist die tatsächliche Studienaufnahme, 5. Schwelle bezeichnet schließlich den Erwerb eines Hochschulabschlusses (vgl. Bundesministerium für Bildung und Forschung, S.65f)

lediglich einen Hauptschulabschluss hat (vgl. Ebenda, S.96). Betrachtet man die unterschiedlichen Chancen im Zeitverlauf ist festzustellen, dass diese sich zwar geringfügig angenähert haben, das aber mehr damit zusammenhängt, dass Kinder aus bildungsnahen Schichten seltener als früher ein Studium aufnehmen, die

Bild 29 Beteiligung der 17- bis 18-Jährigen an der Sekundarstufe II nach Bildungsabschluss und beruflicher Stellung des Vaters in % (1996-2007)¹



Möglichkeiten für Kinder bildungsfernen Schichten dagegen bleibt eher konstant (Abbildung Quelle: Bundesministerium für Bildung und Forschung, 2014).

Es ist ersichtlich, dass die Chancen auf eine höheren Bildungsabschluss mit der sozialen Herkunft (Bildungsniveau der Eltern, Migrationshintergrund etc.) zusammenhängen, also mit dem soziokulturellen Kapital der Herkunftsfamilie.

Das Thema **faire** Leistungsmessung und Benotung muss deshalb in unserer „Leistungsgesellschaft“ eine zentralere Rolle spielen. Bereits ab dem Kindesalter ist ein Leistungsstreben sowie Leistungsdruck von außen feststellbar. Die Leistung gilt gleichzeitig auch als Ordnungsprinzip der Gesellschaft. In Deutschland ist allerdings nach wie vor auch bei guter Leistung kaum Chancengleichheit vorhanden ist.

Etwa die Hälfte aller schichtspezifischen Vorteile und Handicaps gehen von den Eltern auf die Kinder über. Das wiederum führt zu großer Frustration der Schüler unterer Schichten.

Vor allem die Noten entscheiden darüber, wer in welchem Maße Zugang zu weiterer Bildung erhält. Noten dienen zum einen der Kontrolle der pädagogischen Qualität von Lehrplänen, Schularten

oder Schulsystemen, Unterricht und Lehrenden. Des Weiteren sollen Noten aber auch Informationen für Eltern, Schüler und spätere Arbeitgeber über den erreichten Lernstand und verschiedene Fertigkeiten der Schüler geben. Die Gerechtigkeit bei der Vergabe von Noten und ihre Aussagekraft bleiben jedoch umstritten.

Trotz Noten und Bewertungsmaßstäben urteilen Lehrkräfte nicht immer objektiv und haben bei der Bewertung mit verschiedenen psychologischen Einflüssen zu kämpfen, z.B. aufgrund Rückschlüsse durch familiäre Hintergründe der Kinder oder persönlicher Sympathien. Deswegen sind Schülerleistungen oftmals nicht objektiv oder wertfrei.

Um die Chancengleichheit auf Bildung und damit auf sozialen Aufstieg für alle Kinder zu erhöhen, ist es unumgänglich die Noten, welche über den weiteren Bildungsverlauf entscheiden, so objektiv wie möglich zu erfassen.

Ein einfacher Schulleistungstest, sei es eine Klausur, Schulaufgabe, Stegreifaufgabe, Extemporale, ist im Endeffekt nichts anderes als ein quantitativer Fragebogen. Jedoch wird jeder forschungsbezogene Fragebogen hinsichtlich der verschiedenen Gütekriterien erarbeitet und mehrfach überprüft. Der Fragebogen in der Schule, welcher über die Zukunft eines Menschen tatsächlich mitentscheidet, jedoch überhaupt nicht oder nur in sehr geringem Maße.

Gerade die beiden Kriterien Validität und Objektivität können aber leicht in jedem Schultest erhöht werden.

Die Objektivität ist vor allem bei offenen Aufgabenformaten wie das Schreiben von Aufsätzen oder der Bewertung von Übersetzungen bzw. Mediationen. Verschiedene Studien zeigen, dass die Benotung bei solchen Aufgabenstellungen bei verschiedenen Korrektoren zwischen den Noten eins und sechs schwankt. Hierbei spielen nicht nur die unterschiedlichen Bewertungsmaßstäbe an das abgegebene Werk selbst sondern auch subjektive Einflüsse wie der Vorname des Kindes oder das Wissen über die berufliche Tätigkeit der Eltern der Schüler eine Rolle, wie Studien von 1965 bis heute immer wieder belegen.

Derartige Unterschiede konnten aber nicht nur in sprachlichen Fächern sondern auch in den Naturwissenschaften festgestellt werden, obwohl hier automatisch von einer höheren Vergleichbarkeit und Objektivität ausgegangen werden könnte. Die folgenden Ausführungen und Beispiele entstammen den sprachlichen Fächern, vorwiegend dem Fach Englisch.

Wie eigene kleinere Untersuchungen in Seminaren der Englischdidaktik jedes Semester erneut zeigten, sind sich LehramtsanwärterInnen bis kurz vor dem ersten Staatsexamen bei der

Bewertung von Schüleraufsätzen unsicher. Sie können oftmals selbst nicht mit Sicherheit sagen, ob ein von ihnen vermuteter Fehler auch tatsächlich ein grammatikalischer Fehler ist oder die im Aufsatz verwendete Form seine Berechtigung hat. Überraschenderweise würde der Großteil der angehenden Lehrkräfte den vermeintlichen Fehler dann auch als Fehler werten, obwohl sie sich selbst nicht sicher sind. Hier stellt sich die einfache und logische Frage: Warum sollten Schüler, z.B. der 8. Klasse (aus welcher die Probeaufsätze stammten) erkennen, dass es sich bei der verwendeten Struktur um einen Fehler handelt, wenn 10 Studierende der Anglistik sich nicht abschließend einige werden können ob es falsch ist oder nicht?

Bewertungsmaßstäbe müssen also, um die Objektivität zu erhöhen vorab festgelegt werden. Diese können im Nachhinein nur abgeändert werden, wenn sich Probleme der Validität bei der abschließenden Auseinandersetzung mit den Ergebnissen ergeben, Näheres wird im Späteren erläutert. Es ist zu empfehlen, wo möglich, die Maßstäbe in Absprache mit der Klasse vor der Arbeit abzustimmen oder in der Aufgabenbeschreibung festzuhalten, worauf und in welchem Maße bei der späteren Beurteilung Wert gelegt wird.

Schulnoten dienen dazu, den tatsächlichen Leistungsstand der SchülerInnen festzustellen. Bezugsnorm sollte deshalb individuell sein, d.h. die tatsächliche Leistung und nicht die Leistung in Abhängigkeit von den Klassenkameraden. SchülerInnen in besonders leistungsstarken Klassen erhalten tendenziell schlechtere Noten als die selben SchülerInnen in einer leistungsschwachen Klassen erzielen würden, vor allem wenn die Gaußsche Normalverteilung bei der Zuteilung von Noten eine Rolle spielt.

Probleme der Validität ergeben sich bereits vorher, nämlich im Zuge der Testerstellung selbst. Validität bezieht sich auf die Frage: wird tatsächlich gemessen, was gemessen werden soll?

Hierbei gibt es verschiedene Möglichkeiten, warum ein Schultest nicht das misst, was er tatsächlich erheben möchte, nämlich den Leistungsstand der SchülerInnen auf einem bestimmten Gebiet.

Ursachen für mangelnde Validität können u.a. sein:

- Ungenaue oder verwirrende Arbeitsanweisungen
- Zu hohe Ratewahrscheinlichkeit
- Hinweise auf die richtige Lösung durch Fragestellung oder Antwortmöglichkeiten
- Falscher Input

Wichtig ist bei der Erstellung der Aufgaben, klare und eindeutige Fragestellungen bzw. Arbeitsaufträge zu finden. Oftmals scheitern die

Kinder nicht am Lerninhalt sondern an der Aufgabe selbst, da diese schlichtweg falsch verstanden wurde. Ein einfaches Beispiel:

In der Arbeitsanweisung findet sich folgender Satz: *Kreuze die richtige Antwort an.* Diese Formulierung schließt die Richtigkeit mehrerer Antwortvorgaben aus. Selbst wenn ein Kind meint, mehrere Antworten wären richtig, würde es oftmals trotzdem nur eine Antwort geben, da es durch die Arbeitsanweisung verunsichert wäre. Ein gezieltes „Fallenstellen“ sollte in jedem Fall vermieden werden.

Bei verschiedenen Aufgabenformaten besteht ein erhöhtes Raterisiko. Z.B. bei Aufgabenstellungen im Leseverstehen (Ist diese Aussage im Text enthalten? Ja- Nein / Ist die Aussage richtig oder falsch?) oder auch im Hörverstehen bei „same-different“ Aufgaben (sad – sat / bed – bed / meat – meet; Sind die Wörter gleich oder unterschiedlich?). Bei allen Aufgaben ergibt sich eine 50% Ratewahrscheinlichkeit. Für die Aussagekraft der Aufgabe bedeutet dies, es kann nicht mit Sicherheit gesagt werden, wie viel % tatsächliches Fachwissen und wie viel % Rateglück das Punkteergebnis beinhaltet. Die Ratewahrscheinlichkeit kann zumindest verringert werden, wenn eine dritte Möglichkeit miteinbezogen wird. Im Leseverstehen z.B. direkt im Text enthalten, indirekt im Text enthalten, nicht im Text enthalten (allerdings erst ab einem gewissen Alter anwendbar) und im Hörverstehen z.B. too – to-too. Bei Multiple Choice Fragen mit vier Antwortmöglichkeiten liegt die Ratewahrscheinlichkeit immer noch bei 25%. Um diese zu Verringern könnte hier angedacht werden, falsche Antworten negativ zu bepunkten. Die Hemmschwelle einfach zu raten würde entsprechend steigen.

Wenn Multiple Choice Aufgaben z.B. im Vokabeltest, eingesetzt werden, müssen folgende Arbeitsschritte gewissenhaft durchgeführt werden:

1. Vokabel aussuchen, die getestet werden sollen
2. Stem ausarbeiten
3. Distractors finden
4. Eine klare, einfache, verständliche Arbeitsanweisung finden

Hierbei entstehen verschiedene Fehlerquellen, die dazu führen, dass die Schülerwelche anhand von Beispielen im Folgenden dargestellt werden (Madsen, 1983):

- Die Schüler sollen die Antwort nicht allein aufgrund der verwendeten Wortform erkennen können:
She had to help the _____ old man up the stairs.
weak slowly try wisdom

- Nichts durch grammatikalische Hinweise verraten:
She needs to get up earlier so she's buying an ____ clock
time alarm watch bell
- Wörter sollten alle der selben Materie oder einem Themengebiet entstammen:
a small branch =
twig
frog
doom
plum
- Die Antwortmöglichkeiten sollten in etwa den gleichen Schwierigkeitsgrad aufweisen:
They needed lots of training to operate such ____ equipment.
easy sophisticated blue wise
- Alle Antwortvorgaben sollten in etwa gleich lang oder paarweise gleich lang sein.
- Keine Synonyme als Antwortvorgaben verwenden:
His remorse was great indeed.
wealth sadness strength power

Weiterhin ist die Frage zu stellen, welche Kompetenz abgefragt werden soll. Soll das Hörverstehen beispielsweise getestet werden, ist es nicht ratsam, einen zu langen Text zum Hören vorzugeben und danach zu viele Fragen beantworten zu lassen. Hier stellt sich dann nämlich die Frage, inwieweit das tatsächliche Hörverstehen und nicht eigentlich die Merkfähigkeit der Kinder. Auch beim Leseverstehen kann es durch falsche Fragestellungen zu Missverständnissen kommen. Ein Beispiel aus der Grundschule. Ein DinA4 Text über Gorillas, die Aufgabe: Wie lernt ein Gorillababy laufen? Ordne die drei Bilder in der richtigen Reihenfolge.

Abgebildet sind drei Bilder: ein Gorillababy das an der Hand der Mutter läuft, eines das getragen wird, eines das alleine läuft.

Die Antwort eine Kindes: Es geht an der Hand, es läuft alleine, es wird getragen.

Auf Nachfrage erklärte das Kind: „Im Text stand, dass das Gorillababy erst an der Hand der Mutter das Laufen lernt, bis es dann irgendwann alleine laufen kann und wenn es dann vom vielen Laufen müde ist, trägt es die Mama nach Hause.“

Hätte das Kind den Text überhaupt nicht gelesen, wäre die Antwort in der richtigen Reihenfolge angegeben worden, denn es ist logisch, auch bereits für Grundschul Kinder, das erst getragen wird, dann an der Hand gelaufen und dann alleine. Auf Basis des Textes kam es aber zu einem logischen Fehlschluss. Hier ist wieder die Frage, was wurde tatsächlich getestet? Der Input war also an dieser Stelle falsch gewählt, weil verwirrend und die Frage dazu wäre auch ohne den Kontext richtig zu beantworten gewesen.

Allgemein ist festzuhalten, dass es bei einem Test sinnvoll ist, mehrere Aufgaben bzw. Fragen zu einem Gebiet zu stellen um sicher zugehen, dass nicht zufällig eine Frage richtig bzw. falsch beantwortet wurde. In einem Fragebogen würde man so einem komplexen Begriff auch nicht nur eine Frage generieren sondern einen Index aus Mehreren. Das gleiche ist bei Schulaufgaben zu tun. Je mehr Fragen zu einem Aufgabengebiet desto valider ist der Test. Es kann allerdings trotzdem immer wieder passieren, dass einzelne Fragen trotz genauer Vorüberlegungen zu einfach oder zu schwer für die SchülerInnen waren. Dies lässt sich im Nachhinein leicht statistisch überprüfen. Als einfachste Variante bietet sich eine Analyse nach Cronbach's Alpha oder eine einfache Faktorenanalyse an, welche Aufschluss darüber geben, ob die einzelnen Fragen kohärent sind oder nicht. Fallen hierbei einzelne Fragen oder Aufgaben aus dem Raster, können diese nachträglich in der Bepunktung oder Benotung anders gewichtet werden oder gar ganz entfallen. Würden sie das Ergebnis zum Nachteil der SchülerInnen „verfälschen“ sollten sie nicht zur Benotung herangezogen werden.

Insgesamt ist festzustellen, dass sich die Bewertung von Schülerleistungen mit einfachen Mitteln gerechter gestalten lässt. Leider konnten an dieser Stelle nur wenige Beispiele dargestellt werden, welche aber dennoch darauf hinweisen, dass es eine Vielzahl an Fehlerquellen gibt und eine Vielzahl an Möglichkeiten diese zu verhindern.

Literaturverzeichnis

Böttcher, Wolfgang (2005). Soziale Auslese und Bildungsreform. In: APUZ 12/2005. Bpb.

BMBF (Hrsg.) (2006). Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2003. 18. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS. Berlin.

[http://www.sozialerhebung.de/archiv/soz_18_haupt.\[09.07.2014\]](http://www.sozialerhebung.de/archiv/soz_18_haupt.[09.07.2014])

Madsen, Harold S. (1983). *techniques in testing*. Oxford University Press. Oxford.

Statistisches Bundesamt (Hrsg.). *Datenreport 2006*. Wiesbaden (Teil I, Kap.2 (Bildung). Teil II, Kap. 5 (Bildung, Chancen(un)gleichheit bei Bildung) unter: www.destatis.de [09.07.2014]

Vester, Michael (2004). *Die Illusion der Bildungsexpansion. Bildungsöffnungen und soziale Segregation in der Bundesrepublik Deutschland*. In: Engler, Steffani / Kraus Beate (Hrsg.) (2004). *Das kulturelle Kapital und die Macht der Klassenstrukturen*. Juventa. Weinheim und München.