

Am Beispiel der Logopädie-Ausbildung

JÜRGEN KOHLER

E i n l e i t u n g

Im letzten L.O.G.O.S. interdisziplinär

lasen Sie den theoretischen Teil der

„Leistungsbewertung am Beispiel

der Logopädie-Ausbildung“. Hier

folgt nun ein Erfahrungsbericht.

Wir haben an der Lehranstalt

für Logopädie Reichenau

die Anwendung der Taxo-

nomiestufen ausprobiert,

allerdings zunächst nur

für den schriftlichen Teil

der Prüfung im logopädi-

schen Teilgebiet Aphasie.

Nachfolgend wird der chrono-

logische Verlauf dieses Ver-

suchs wiedergegeben.

Schlüsselwörter:

Logopädie-Prüfung,

Taxonomiestufen,

Testgütekriterien,

Test-Theorien



LEISTUNGSBEWERTUNG

EMPIRISCHER TEIL

CHRONOLOGISCHER PROJEKTVERLAUF

Leistungsbewertung der schriftlichen Prüfung im logopädischen Teilgebiet Aphasie

1. Veröffentlichung des theoretischen Teils der „*Leistungsbewertung am Beispiel der Logopädie-Ausbildung*“. Die Studierenden beschäftigen sich mit diesem theoretischen Teil
2. Diskussionen zur Klärung bzw. Vertiefung der Sachverhalte
3. Eine Klausur mit Erwartungshorizont wird von Lehrlogopäden nach Maßgabe der Taxonomiestufen konstruiert
4. Die Klausur wird von den Studierenden des ersten Kurses geschrieben und gleich danach mit ihnen mündlich besprochen
5. Verschiedene Bewerter geben unabhängig voneinander ein Urteil ab:
Die Dozenten und die Studierenden korrigieren alle Klausuren nach Maßgabe des Erwartungshorizontes. Die Korrektur beinhaltet die
 - Punktvorgabe, plus
 - Transformation in eine Notenskala und
 - die Zuordnung der Aufgaben zu den Taxonomiestufen Blooms
6. Erfahrungen bei der Bewertung werden ausgetauscht und diskutiert sowie die Ergebnisse verglichen
7. Versuch, andere Dozenten für die Problematik der Leistungsbewertung zu sensibilisieren, besonders den Wert der Taxonomiestufen auf andere Fächer zu übertragen. Beobachtung der Auseinandersetzung mit dem Thema „Leistungsbewertung“ im Schulalltag.

Erklärtes Projektziel war **Offenheit** und **Transparenz** im Bewertungsprozess zu erreichen sowie die **Gütekriterien der Prüfungsaufgaben zu erfassen bzw. zu verbessern**.

Zunächst wurde in Kurs 1 das Projekt wie oben beschrieben chronologisch durchlaufen.

Die Stoffmenge wurde den Studierenden so genau

wie möglich angegeben. Die Dozenten gaben Literatur an und verwiesen auf die entsprechenden Unterrichtseinheiten. Um die Unabhängigkeit der Bewertungen zu sichern, kopierten wir die Originale der verfassten Klausuren, so daß jeder Bewerter zur Korrektur eine Kopie ohne Randnotizen von anderen Beurteilern erhielt. Absprachen zwischen den Beurteilern fanden während der Korrekturphase nicht statt.

Ich will Ihnen an drei Beispielen zeigen, welche Erfahrungen wir gesammelt und diskutiert haben. Diese Erfahrungen sollen anregen, konkrete Konsequenzen der Taxonomien für die Leistungsbewertung kritisch zu diskutieren. Die Beispiele sollen besonders die Schwierigkeiten verdeutlichen, auf die wir gestoßen sind, jedoch ist dies nur eine kleine Auswahl an denkbaren Problemen. Wer sich für die allgemeinen zahlreichen Fragen der Leistungsdiagnostik interessiert, kann dies an anderer Stelle nachlesen (z.B. Ingenkamp 1971). Ich erhoffe mir durch die beispielhafte Darstellung unserer Probleme und Erfahrungen mehr Anschaulichkeit und Bezug zur Logopädieausbildung für die Leserschaft.

Erste Hinweise auf Probleme der Aufgabenstellung ergaben sich bei der mündlichen Besprechung derselben, die kurz nach der Klausur stattfand.



Das ganze Ausmaß an Problemen wurde aber erst durch die Korrekturarbeit deutlich.

Erstes Beispiel: Fehler bei der Aufgabenkonstruktion

Besonders solche Aufgaben, die in die Taxonomiestufe *Anwendung* fallen, aber auch *Analyse-* oder *Synthese-*Aufgaben, bedürfen wohl dringend eines Probedurchlaufes, bevor sie tatsächlich zur Leistungsmessung verwandt werden. Bei der Aufgabenkonstruktion können sich Fehler einschleichen, die erst beim Korrekturvorgang bemerkt werden. Die Aufgabe an sich kann mißverständlich sein oder der Erwartungshorizont ist fehlerhaft.

Hierzu ein Beispiel:

Lutz (1992) integriert in ihrem Ansatz verschiedene schon existierende andere Ansätze. Sie verwendet Elemente der Deblockierungsmethode (Weigl 1979), bezieht sich auf den linguistischen Ansatz (Kotten 1984) und erwähnt auch PACE (z.B. Glinde-mann / Springer 1989) als Quelle ihrer eigenen Arbeit. Die Studierenden mußten nun ihr vorhandenes Wissen zu den oben erwähnten Ansätzen anwenden und dem für sie unbekanntem MODAK-Modell zuordnen. Dazu lag ihnen eine verkürzte Handlungsbeschreibung der sechs MODAK-Phasen vor.

Hier ein Auszug:

Phase 1:

Zeigen der Bilder: Vier Bilder liegen auf dem Tisch. Der Therapeut nennt einen Satz, der Patient zeigt das entsprechende Bild. Wenn er nicht zeigen kann (weil er den Satz nicht versteht), zeigt der Therapeut mit ihm gemeinsam.

Die erste Aufgabe war, die sogenannten Sprachfunktionen aus der Deblockierungsmethode den einzelnen Phasen zuzuordnen (Ich setze die Kenntnis der Sprachfunktionen beim Leser voraus).

Bei der Korrektur wunderte ich mich,

daß einige Studierende Sprachfunktionen angaben, die „überhaupt nicht paßten“. Warum schreibt jemand „mündliches Benennen“ bei Phase 1, wenn der Patient doch nur „Bilder zeigen“ muß. Die letzte von mir korrigierte Klausur brachte die Entwunderung bzw. die Verwunderung über mich selbst.

Frau E. wählte zur Darstellung ihrer Lösung eine Tabellenform, wobei sie die Sprachfunktionen getrennt einmal dem Patienten und zum anderen dem Therapeuten zuordnete. Während ich die Therapiesituation nicht interaktionistisch betrachtet hatte, waren einige Studierende völlig selbstverständlich davon ausgegangen, daß in jeder Phase Patient und Therapeut sprachlich handeln. Es ist also absolut richtig, auch dem Therapeuten die Sprachfunktionen zuzuordnen und nicht nur dem Patienten. Da predigt man im Unterricht den Interaktionismus als die angemessene Perspektive zur Betrachtung aphasischer Kommunikation und dann so was! Manchmal entdeckten wir solche Diskrepanzen zwischen Aufgabenstellung und Erwartungshorizont schon früh (nicht erst bei der letzten Klausur) und gerieten in ein Dilemma, das man so beschreiben kann:

Wir hatten mit der Konstruktion der Aufgaben und der Festlegung des Erwartungshorizontes A gesagt.

Müssen wir mit der Anwendung des Erwartungshorizontes nun B sagen, obwohl wir erkannten, daß A einen Fehler hat?

Es war klar, daß die Gerechtigkeit bei strikter Anwendung des Erwartungshorizontes nicht gegeben war. Bei Unstimmigkeiten zwischen Aufgabe und Erwartungshorizont muß entweder die Aufgabe eliminiert oder der Erwartungshorizont korrigiert werden. Es ist nichts Besonderes, daß Tests einiger Probedurchläufe bedürfen, um deren Aufgaben zu validieren. Insofern befinden wir uns in bester Gesellschaft mit allen standardisierten Testverfahren der Diagnostik. Doch wir mußten uns an dieser Stelle fragen, ob dieser Aufwand zu schaffen ist und waren uns einig, daß dies ziemlich unwahrscheinlich ist.

Zweites Beispiel: Gewichtungsprobleme oder: was ist eigentlich wichtig?

Das zweite Beispiel soll angeführt werden, um das Problem der Gewichtung darzustellen. Innerhalb der Literatur rund um die Aphasietherapie werden die Begriffe „Deblockierung“ und „Stimulierung“ uneinheitlich gebraucht (vgl. Kotten 1989, S.85). Springer (1979, S. 469) schrieb bestimmten Alltagssituationen einen deblockierenden und stimulierenden Effekt zu, unabhängig von der „reinen“ Deblockierungsmethode. Die Studierenden, die noch keinerlei Erfahrung in der Aphasiebehandlung hatten, sollten Vermutungen darüber anstellen, welche charakteristischen Merkmale solche Alltagssituationen haben könnten.

Diese Aufgabe erfordert in gewissem Maße Kreativität (Synthese), da man sich konkrete Situationen vorstellen muß. Der Erwartungshorizont bleibt relativ offen, da vielerlei Situationen stimulierend sein können. Es gibt auch kein absolutes richtig oder falsch, denn die Reaktion auf eine Situation muß nicht für alle Aphasiker die gleiche sein. Die Nennung von charakteristischen Merkmalen ist zudem eine Abstraktionsleistung, da die konkreten Situationen nach ihren Gemeinsamkeiten geordnet werden müssen. Hierzu ein Beispiel: *Einen Ausflug ins Grüne machen, von den Enkelkindern Besuch bekommen oder beim Kartenspiel gewinnen* sind ganz unterschiedliche Situationen, die als gemeinsames charakteristisches Merkmal „Freude“ haben können, was zur sprachlichen Stimulation beitragen könnte.

Die potentielle Vielfalt der Antworten macht die Punktevergabe schwer. Honoriere ich die Fülle der gebrachten Beispiele? Soll ich die genannten charakteristischen Merkmale mehr belohnen, die auch in der Literatur häufig genannt werden oder zählt der ungewöhnliche, vielleicht etwas abenteuerliche Einfall von jemandem mehr, der noch nie mit Aphasikern Kontakt hatte? Neben den spezifischen Schwierigkeiten der Punkteverteilung dieser Aufgabe gab es auch allge-

meine Gewichtungsprobleme zwischen den Aufgaben:

Warum gebe ich für diese Aufgabe im Vergleich zu anderen, die Wissen oder Analyse verlangen, weniger Punkte, obwohl die Synthese eigentlich die höhere Leistung innerhalb der Taxonomiestufen ist?

Was ist eigentlich wichtig: Wissen haben, anwenden können, kreativ sein oder ...? Darauf will ich später noch einmal eingehen. Ein qualitativ durchaus erwünschter Fördereffekt sei hier aber vorweggenommen. Es zeigte sich, daß nicht nur die Komplexität der Leistung zunimmt, je höher die Taxonomiestufe, sondern schon allein durch den Spiegel der Taxonomiestufen auch das Bewußtsein der Studierenden wächst für die Komplexität der realen Situationen, in denen Therapie stattfinden kann. Problematisch dabei ist, daß sich die Richtigkeit der Antworten im Grunde erst durch die praktische Erprobung der Ideen am realen Therapieschauplatz erweist, was natürlich im Rahmen einer schriftlichen Klausur schwerlich gegeben ist.

Drittes Beispiel: Transformationsfehler

Mit Transformationsfehlern sind solche gemeint, die aufgrund der Umrechnung von einem Bewertungssystem in ein anderes entstehen.

Bei der Punktevergabe hatten wir insgesamt fehlerbereinigte 50 Punkte zu vergeben. Nun könnte man mehr oder weniger willkürlich entscheiden, daß jeder Punkt ein Zehntel Abzug bei der Umrechnung in das traditionelle Notensystem darstellt. 49 Punkte entsprächen dann der Note 1,1. Wir haben uns aber dafür entschieden, eine gebräuchliche Umrechnungstabelle zu benutzen, nämlich die des Landes Rheinland-Pfalz, die auf 100 Punkte ausgelegt ist. Dafür haben wir unsere Punkte einfach verdoppelt und dann die vorgegebene Umrechnungstabelle benutzt.

Die beiden Bewertungssysteme sind aber nicht kongruent, und dadurch gibt es in gewissen Bereichen Verzerrungen. Dies sei an folgendem Beispiel verdeutlicht. (Tab. 1)

Der Unterschied zwischen den Bewertern	
für Klausur A betrug	91,5-86,5 = 5,0 Punkte
für Klausur B	77-71,5 = 5,5 Punkte
jedoch war der daraus resultierende Notenunterschied	
bei A	1,4-1,9 = -0,5
bei B	3,9-4,1 = -0,2

Tab. 1

In Worten ausgedrückt passiert folgendes: Ein „kleiner“ Leistungsunterschied führt zu einem größeren Bewertungsunterschied als ein „großer“ Leistungsunterschied.

Weiterführende Diskussion: Quantitative und qualitative Ergebnisse

Ein Ergebnis des Projekts entstand durch die **Zuordnung der Aufgaben zu den Taxonomiestufen**. Jeder Bewerter mußte entscheiden, welche Aufgabe zu welcher Taxonomiestufe gehört. Die Korrelation zwischen den Bewertern würde eine Vermutung darüber erlauben, inwiefern die Aufgaben tatsächlich eine Stufe prüfen. Das Ergebnis war erwartungsgemäß bescheiden. Wir konnten nur bei zwei von acht Aufgaben eine große Übereinstimmung zwischen den Bewertern feststellen. Alle anderen Aufgaben wurden nicht eindeutig zugeordnet. Die unbefriedigende Trennschärfe zwischen den Stufen war somit bestätigt worden. Wir glauben jedoch, daß die Idee aller Bloomschen Taxonomiestufen ihren Wert behält. Besonders die Kategorien *Anwendung* und *Synthese* bzw. *Evaluation* erscheinen uns als wertvolle Leitlinien zur Beschreibung von ausbildungsrelevanten Leistungen, auch wenn die eindeutige Kategorisierung bzw. quantitative Messung derselben bei unserer schriftlichen Prüfung schwierig war und dies auch im allgemeinen zu sein scheint. Die Tatsache, daß etwas schwierig kategorisiert und gemessen werden kann, ist noch kein Grund, seinen Wert anzuzweifeln. Schade wäre es, wenn solche Leistungen deswegen nicht Gegenstand einer Leistungsbeurteilung sein würden und stattdes-

sen nur solche Aufgaben zum Tragen kommen würden, die ihrer leichten Meßbarkeit wegen „gerecht“ sind. Das bloße Abfragen von Wissen ist sehr leicht meßbar, doch würde das der Qualität der Leistung (=Validität) und damit der Ausbildung

einen Bärendienst erweisen. Ich halte es im Gegenteil für sinnvoll, eine Eingrenzung des prüfungsrelevanten Wissens vorzunehmen, damit die kognitiven Ressourcen der Studierenden bei der Prüfungsvorbereitung nicht durch bloßes Auswendiglernen überstrapaziert werden. Wissen, das man nur für die Prüfung paukt, um es danach ganz schnell wieder zu vergessen, kann man im Zweifelsfall nachschlagen. Weniger kann hier mehr sein. Die höheren Stufen der Taxonomie verlangen eher Übung als Pauken und zwar relativ unabhängig von der Stoffmenge. Allerdings ist die Frage *was ist wichtig?* auch hier brisant. Die Entscheidung darüber, was unbedingt gewußt werden sollte, nimmt dem Prüfer niemand ab.

Trotz aller Probleme, die oben ja nur beispielhaft erläutert wurden, war die Korrelation der Beurteilungen zwischen den zwei unabhängig voneinander bewertenden Lehrlogopäden mit 0,91 gut. (Abb. 1, S. 185) Die **Übereinstimmung zwischen zwei Beurteilern** ist, wie weiter oben erläutert, ein Maß für die Objektivität einer Messung. Der Zahlenwert 0,91 ist ein statistisch ermittelter Kennwert für die Übereinstimmung, wobei 0,00 maximale Differenz und 1,00 maximale Deckung in der Beurteilung bedeuten würden. Man kann in diesem Fall also von einer guten Übereinstimmung sprechen.

Die qualitativen Ergebnisse des Projekts waren für uns weitaus interessanter als die quantitativen. Diese Ergebnisse werde ich in zwei Schritten darlegen, die dem chronologischen Projektverlauf Punkt 6 und 7 entsprechen. Zuerst werden die Ergebnisse kurz nach der Klausur dargestellt (vgl. Punkt 6.) und danach die weitere Auseinandersetzung rund um das Thema Leistungsbewertung im Schulalltag sowie der Versuch, den Wert der Taxo-

nomiestufen auf andere Fächer zu übertragen (vgl. Punkt 7.).

Qualitative Ergebnisse kurz nach der Klausur - kurzfristige Effekte

Zunächst bekamen wir das Feedback der Studierenden, daß die Art der Aufgabenstellung dem Thema Aphasie-therapie dienlich war. Auch wenn die Notenvergabe fehlerhaft und unpräzise war, so erfuhren die theoretischen Begriffe mehr Präzision. Das schloß die Diskussion um die Uneindeutigkeit der theoretischen Terminologie mit ein (Ist die (De-)Blockierung allgegenwärtig?). Somit hatte die Klausur die erhoffte Förderfunktion innerhalb der Ausbildung im logopädischen Teilgebiet Aphasie geleistet. Nicht zuletzt hatten die Lehrlogopäden etwas gelernt (von wegen Interaktion!).

Weitaus wichtiger war uns jedoch der Effekt auf einer übergreifenden Ebene. Durch die Veröffentlichung der zahlreichen Schwierigkeiten erfuhren viele der Beteiligten ein angenehmes Gefühl der Gelassenheit gegenüber dem großen Anspruch der Gerechtigkeit bei der Notenvergabe. Die Mühe der Lehrlogopäden bei der Projektplanung und -durchführung hatte sich insofern gelohnt, als klar wurde, daß wir mit unseren bescheidenen Mitteln an zeitlichen und methodischen Ressourcen immer Fehler machen werden. Es wäre auch vermessen gewesen, etwas anderes zu erwarten, nachdem sich schon seit Generationen hochkarätige Erziehungswissenschaftler die Köpfe zur schulischen Leistungsmessung zerbrochen hatten, viele Fehlerquellen zwar entdeckten, aber weitaus weniger zu ihrer Beseitigung beitragen konnten (Ingenkamp 1971). Die Transparenz der Fehler wirkte als vertrauensbildende Maßnahme und sicherte uns das Mitgefühl der Studierenden bei der schwierigen Aufgabe der Leistungsbewertung. Die Maßnahme, daß die Studierenden ihre eigene Arbeit korrigieren mußten, führte jedem Einzelnen die vielseitigen Probleme der Bewertung

vor Augen. Das eigentlich Wichtige war das gegenseitige Verständnis. Offenheit war das wesentliche Bestimmungsmerkmal von empfundener Gerechtigkeit.

Qualitative Ergebnisse im Schulalltag - mittel- und langfristige Effekte

Zunächst stelle ich die weiteren Maßnahmen vor, die ergriffen wurden, um auch andere an der Logopädieausbildung beteiligte Dozenten für die Problematik der Leistungsbewertung zu sensibilisieren. Dies geschah über die Veröffentlichung des theoretischen Teils sowie über die beispielhafte Erläuterung des Klausurprojekts im Rahmen einer Dozentenkonferenz. Das Interesse für diese Problematik und die Auseinandersetzung mit dem Thema war je nach Dozent sehr unterschiedlich. Einige fühlten sich nicht betroffen, da sie kein Prüfungsfach unterrichten. Andere sagten ganz offen, daß sie im Rahmen ihrer Honorartätigkeit keine Zeit für die notwendige intensive Auseinandersetzung hätten. Wieder andere stellten die Übertragbarkeit der Taxonomien wegen der Spezifität ihres Faches in Frage. Z.B. äußerte der Anatomiedozent Zweifel darüber, ob die höheren Stufen für sein Fach relevant seien. Hier käme es vor allem auf Wissensvermittlung an. Es gibt aber auch Dozenten, die sich bei ihrer Prüfungstätigkeit ganz ausdrücklich auf die Taxonomiestufen beziehen.

Während die quantitativen Ergebnisse kurz nach der Klausur durch eine Besprechung mit dem gesamten Kurs 1 sowie vieler Einzelinterviews systematisch erfaßt wurden, liegen für die qualitativen Effekte des weiteren Verlaufes keine Fragebogen- oder Interviewdaten vor. Aus diesem Grund kommen drei unterschiedliche Perspektiven zu Wort. Eine Schülerin des ersten Kurses, eine Schülerin des zweiten Kurses sowie die am Projekt hauptsächlich beteiligte Lehrlogopädin schildern in eigenen Worten diesen Prozeß. Es beginnt die Schülerin des zweiten Kurses. Bei diesem Kurs sollte das gesamte Prozedere (vgl.

chronologischer Projektverlauf) wiederholt werden. Wir ließen dieselbe Klausur schreiben, hatten aber die aus der Erfahrung mit Kurs 1 hervorgegangenen Fehler bzw. Unstimmigkeiten ausgemerzt und erhofften uns im Ergebnis bzw. im weiteren Projektver-

lauf, Effekte beobachten zu können, die auf die Verbesserung der ersten Klausurversion zurückzuführen sind (= experimentelles Setting). Doch es kam alles ganz anders. Lesen Sie nun den Bericht von Sabine Wieser, die beschreibt, was geschah:

Kleines Mißverständnis mit großen Folgen

Nachdem auch in Kurs 2 in den ersten Monaten der Logopädieausbildung das Teilgebiet Aphasie intensiv unterrichtet worden war, sollte Ende April in diesem Fach die erste Klausur geschrieben werden. Ca. 6 Wochen vorher erhielten wir ein (15 Seiten umfassendes) Papier, den sogenannten theoretischen Teil der „Leistungsbewertung am Beispiel der Logopädieausbildung“.

Es wurde größtenteils als sehr positiv empfunden, daß sich die Lehrlogopäden so intensiv mit dem Thema Leistungsbewertung beschäftigten. Schließlich ist es ja in unserem Interesse, sich einer möglichst objektiven, gerechten und nachvollziehbaren (transparenten) Leistungsbeurteilung „ausgeliefert“ zu wissen. Nach der Lektüre des ausgeteilten Papiers bestanden aufgrund der wissenschaftlichen Ausrichtung des Projekts (und somit auch des Textes) schon noch einige Unklarheiten, die im Rahmen eines Gesprächs mit dem ganzen Kurs weitgehend auszuräumen versucht wurden. Insbesondere auf die Taxonomie von Lernzielen wurde nochmals näher eingegangen.

Je näher der Klausurtermin herandrückte, desto mehr wurden auch innerhalb des Kurses Anzeichen von Nervosität, Unsicherheit und Angst spürbar. Für einige sollte diese Klausur nach Jahren wieder die erste schriftliche Prüfung überhaupt sein.

Entspricht mein Wissensstand dem, was die Dozenten erwarten? Wird es mir gelingen, im entschei-



denden Moment das schriftlich zu formulieren, was ich gedanklich präsent habe?

Vor allem aber: Werde ich überhaupt die Fragestellung richtig verstehen, d.h. in einer Auf-

gabe diejenige Taxonomiestufe erkennen, auf die es dem Dozenten hierbei ankommt?

Solche und noch viele ähnliche Gedanken drängten sich uns zunehmend auf. Viele fühlten sich unter Leistungsdruck, der jedoch weniger von den Dozenten ausging als von ihnen selbst.

Eine plausible Möglichkeit, die Aufregung vor der Klausur etwas zu mildern, sahen einige Studierende darin, sich mit Kurs 1 kurzzuschließen, der letztes Jahr ja ebenfalls eine Klausur im Fach Aphasie geschrieben hatte. Besonders gute freundschaftliche Kontakte zwischen einzelnen Studierenden beider Kurse brachten ca. eine Woche vor dem Klausurtermin den Erwartungshorizont vom Vorjahr in Umlauf.

Zwar erhielt man dadurch schon einige Anhaltspunkte bezüglich dem, was geprüft wurde, dennoch lag das Hauptinteresse in der Art der **Fragestellung**. Und tatsächlich erfuhren zunächst Einzelne auch in diesem Punkt von Kurs 1 Hilfsbereitschaft. Am Tag vor der Klausur war dann die Situation eingetreten, daß ein paar von uns also auch im Besitz der letztjährigen Klausurfra-

gen waren. Von denjenigen, die sie (noch) nicht besaßen, hatten jedoch einige davon mitbekommen. Um innerhalb des Kurses für Chancengleichheit zu sorgen, wurden dann auch die Fragen für **alle** kopiert. (Diejenigen, die sich um deren „Verfügbarkeit“ bemüht und nachvollziehbarerweise nicht beabsichtigt hatten, daß die Fragen ebenso die Runde machten wie der Erwartungshorizont, waren damit weniger einverstanden). Diese Situation sorgte in der Klasse für eine derart angespannte Atmosphäre, wie ich sie vorher noch nicht erlebt hatte. Ich kann es vielleicht beschreiben als eine Mischung von Mißtrauen, einzelkämpferischem Egoismus und Angst auf der einen, aber auch solidarischem Klassenzusammenhalt auf der anderen Seite. Als dann am nächsten Tag die Klausur ausgeteilt wurde, traf wohl jeden von uns fast der Schlag! Wir hatten (fast) haargenau dieselben Fragen vor uns, wie die, die noch am Vortag an alle verteilt worden waren. Die meisten hatten sie sich selbstverständlich auch noch mehr oder weniger gründlich (in Zusammenhang mit dem Erwartungshorizont) zu Gemüte geführt, obwohl **niemand** geglaubt hatte, daß wir tatsächlich **dieselbe** Klausur wie Kurs 1 schreiben würden. – Einige von uns fühlten sich wie gelähmt. – Wir konnten es kaum fassen, eine Klausur vor uns zu haben, deren Aufgabenstellungen inklusive Antworten man bereits kannte! So hatte sich das wirklich keine/r von uns vorgestellt! Die 1 1/2-Stunden Bearbeitungszeit gingen vorbei, ohne daß an die Aufsicht führende Dozentin irgendwelche Fragen bezüglich der Aufgabenstellung oder ähnliches gestellt wurden. (Dies war wohl in Kurs 1 in großem Maße geschehen). Daß den Korrektoren die ganze Sache schon nach Durchsicht der ersten paar Arbeiten „spanisch“ vorkommen würde, war wohl den meisten von uns klar.

Normalerweise wären in der anschließenden Pause Freude und Erleichterung darüber, daß man es nun hinter sich hat, angebracht gewesen. So aber zeigte sich nichts als Fassungslosigkeit, Enttäuschung, Entsetzen über die Sachlage. Bereits am Nachmittag war die Besprechung der Klausur angesetzt worden. Entsprechend der Erwartung, daß uns hierbei das „Fiasko“ in vollem Umfang deutlich gemacht werden würde, fiel dieses Gespräch mit den beiden Dozenten dann auch aus. Die Atmosphäre war miserabel. Ein gewisses Vertrauensverhältnis, das bis dahin zwischen den Logopäden und uns entstanden war, schien zerstört. Wir hatten das Gefühl, daß uns von den Dozenten unterstellt wurde, sie auf plumpe Art und Weise bewußt hinters Licht geführt haben zu wollen. Außerdem hatten wir damit die Durchführung des Projekts der Leistungsbewertung hinsichtlich dieser Aphasiaklausur verpatzt. Letzteres war uns allen zweifellos einsichtig. Die von Seiten der Dozenten angedeutete Schuldzuweisung konnten wir jedoch auf keinen Fall unwidersprochen lassen. **Niemals** hatte irgend jemand von uns die Absicht gehabt, die Klausurergebnisse auf unehrliche Art und Weise zu manipulieren oder das Projekt zu unterlaufen. Welchen Grund hätten wir auch dafür haben sollen? Schließlich waren ja hauptsächlich wir selbst die Leidtragenden der ganzen Angelegenheit, weil wir nun natürlich keinerlei Rückmeldung über unseren Leistungsstand bekamen. Worin könnte nun aber die Ursache dieses „Klausur-Fiaskos“ gelegen haben? Wir sehen sie hauptsächlich in einem Mißverständnis zwischen Dozenten und Studierenden. Im nachhinein konnten sich einige von uns daran erinnern, daß einmal gesagt worden war, wir würden „die gleiche“ Klausur

schreiben wie der 1. Kurs. Was aber genau ist unter „die gleiche“ zu verstehen? Wenn es bei einem Test Gruppe A und Gruppe B gibt (gleiches Themengebiet, aber unterschiedliche Fragen von vergleichbarem Schwierigkeitsgrad), kann man dann nicht auch noch sagen, es schreiben alle den gleichen Test? Es wäre wohl alles anders gelaufen, wenn die Durchführung des Projektes eindeutiger gewesen wäre. Durch eine ausdrückliche Formulierung wie beispielsweise „*Ihr werdet (fast) haargenau dieselben Fragen gestellt bekommen wie Kurs 1 letztes Jahr. Des Projekts wegen muß es auch in eurem eigenen Interesse sein, keinerlei Informationen hierüber von den anderen einzuholen*“ hätte dies mit ziemlicher Sicherheit erreicht werden können. In vielen Bereichen (z.B. gymnasiale Oberstufe, verschiedene Studiengänge) ist es eine Selbstverständlichkeit, sich u.a. durch Klausuren der Vorjahrgänge auf die eigene, bevorstehende vorzubereiten. Von daher ist es bestimmt nicht zu verurteilen, daß einige Studierende des 2. Kurses den Versuch unternommen haben, sich die Klausur (mit Erwartungshorizont) ihrer Vorgänger zu besorgen. Daß denjenigen eine Herausgabe derselben an uns eigentlich untersagt worden war, hatten in der nötigen Deutlichkeit dort wohl auch nicht alle verstanden. Als Studierende der Logopädie ist unser aller Anliegen die Kommunikation unter den Menschen. Mißverständnisse können vielleicht als eine Form fehlerhafter Kommunikation gesehen werden. Hat man sie erst mal als solche erkannt, gilt es, sie weitgehend aufzuklären. Daß uns dies in einem zweiten Gespräch mit den Dozenten gelungen ist (was ich persönlich zumindest hoffe), stimmt mich glücklich. Ansonsten wären wir als (angehende) Logopäden und Logopädinnen vielleicht doch fehl am Platz.

Sabine Wieser, 2. Kurs

Ein kleiner Fehler mit weitreichenden Folgen. Wir hatten nicht genügend herausgestellt, daß wir die Klausur im zweiten Kurs wiederholen wollten. Wir hatten zwar enorm viel Zeit aufgewandt, um die edlen Ziele des Projekts zu verdeutlichen und die Kriterien der Bewertung genau zu erläutern, doch den Zusammenhang im experimentellen Setting nur mal nebenbei erwähnt. Die Effekte des experi-

mentellen Settings (= ob die verbesserte Klausur wirklich weniger Auswertungsprobleme mit sich bringt) konnten wir somit nicht beobachten. Auch konnten wir mit dem zweiten Kurs nicht noch einmal den Weg gehen, der die unabänderlichen Schwierigkeiten der Bewertung (z.B. das Gewichtungssystem) transparent gemacht hätte. Mit Kurs 2 sind wir einen anderen Weg gegangen: Da wir ge-

meinsam dieses Mißverständnis klären und eine Enttäuschung überwinden konnten, hatte der Kommunikationsmißerfolg wiederum mehr Verständnis füreinander gebracht, allerdings nicht auf der Ebene der Leistungsbewertung.

Durch die Reflexion von Ann Flügel (Lehrlogopädin) erfahren Sie nun, welche Konsequenzen beide Wege für ihre weitere Prüfungstätigkeit hat:

Erwartungshorizont



Immer wieder wurde ich in der Korrekturphase damit konfrontiert, daß der Erwartungshorizont nicht zur Fragestellung paßte und insofern die Antworten der Schüler auch nicht erwartungsgemäß ausfielen. Der Erwartungshorizont schien zu eng gefaßt, es verbargen sich darin beispielsweise inhaltliche Ungenauigkeiten. Mein Problem: Wie stark klebe ich nun an diesem Horizont? Wieviel Spielraum habe ich, davon abzuweichen? In dieser, meiner ersten Korrektur einer Klausur, entschied ich mich, so genau wie möglich daran zu bleiben und machte mir Notizen, wie ich was bewertete, damit ich es konsequent bei allen durchhalten würde. In der anschließenden Diskussion mit den Schülern über die Fehler des Erwartungshorizontes stellten wir uns z.B. die Frage, ob es nicht überhaupt besser sei, diese Vorgabe wegzulassen, um „freier“ korrigieren zu können. Konsequenz aus diesen Überlegungen für mich war, den Erwartungshorizont zu verändern, um dann die gleiche Aphasiaklausur mit dem zweiten Kurs in der neuen, hoffentlich verbesserten Version zu schreiben. Zunächst versuchte ich, die Aufgabenstellung an den kritischen Punkten zu verdeutlichen und dann den Erwartungshorizont offener mit mehr Spielraum zu formulieren. Z.B. erwähnte ich 5 mögliche Antworten, für die 3 Punkte vergeben werden konnten, nicht mehr wie in der ersten Version 3 feststehende Items, für die es genau 3 Punkte geben kann. Somit hatte ich etwas aus

dem ersten Durchgang gelernt, was ich nun ausprobieren wollte. Meine Erwartung war, daß den Schülern die Eigenbewertung und auch uns Lehrlogopäden die Beurteilung

leichter fallen würde. Außerdem hofften wir, durch diese Verbesserung der Klausurkonstruktion viel Unzufriedenheit, die alle Beurteiler über die Probleme im Umgang mit der Aufgabenstellung und dem Erwartungshorizont geäußert hatten, zu vermeiden. Wenn die Maßgabe zur Korrektur mehr Offenheit und Spielraum zuläßt, scheint die Bewertung leichter zu fallen und mehr individuelle Problemlösungen, die sehr gut sein können, zuzulassen. Meine Idee nach den Erfahrungen: Aufgaben mit Erwartungshorizont zu konstruieren macht Sinn, wenn dieser in seiner Abstimmung nicht zu starr und zu eng gefaßt wird und mehrere mögliche Lösungen für eine bestimmte Punktzahl zuläßt.

Korrekturschwierigkeiten

Der zweite Aspekt, der mich während dieses Projektes beschäftigt hat, ist mein Anspruch, einem hohen Maß an Objektivität genügen zu wollen. Immer wieder wollte ich vergleichbar korrigieren und niemanden bevorzugen. Nach einer weiteren Klausurkorrektur im Fach-

bereich „Kindersprache“ ist mir klar geworden, daß dies besser gelingt, wenn man Frage für Frage, statt Klausur für Klausur korrigiert. So hat man die Lösung zur gleichen Frage von vorherigen Schülern noch im Kopf und kann sich daran ein wenig orientieren.

Dann habe ich versucht, auf meine Stimmungen zu hören. Wer schon einmal einen Berg mit 20 Klausuren vor sich liegen hatte, weiß, wie arbeitsam dieser Job ist. Manchmal war ich einfach müde und hatte keine Lust mehr, dagegen saß ich einmal auf der Terrasse in der Sonne und war sehr fröhlich. Beeinflußt meine Stimmung die Bewertung? Die Vermutung steht im Raum, und ich habe das Wissen um den Effekt, unter gewissen Umständen eher milde oder aber streng zu bewerten, während der Korrektur im Kopf gehabt.

In diesem Zusammenhang gab es auch einen übergeordneten, von den Details der Klausur unabhängigen Milde-Effekt (vgl. Zimbardo, 1983, S.591). Ich dachte oft an meine eigene Ausbildung und an die Bedeutung, die das Fach „Logopädie“ für mich hatte. Das war das Wichtigste, das sollte ja mein Beruf werden, hier wollte ich gut sein. Als ich nun auf „der anderen Seite“ saß, dachte ich: Ich will nicht so streng sein, diese erste Klausur in Logopädie hat doch eine Bedeutung. Dieses Gefühl bestätigte sich, nachdem die Ergebnisse bekannt wurden. Der Notendurchschnitt war tatsächlich eher durchschnittlich, und sowohl

einige Schüler als auch ich waren mit den Resultaten nicht so glücklich. Die anschließende offene Diskussion über die Probleme der Bewertung und die Schwierigkeiten bei der Klausurkonstruktion machten dieses Gefühl dann wieder wett.

Vertrauensaufbau als Ergebnis

Vertrauen zu schaffen gegenüber der Bewertungssituation war ein großes Ziel, das wir erreicht haben. Die Beurteilung von Leistungen schien transparenter, das Eingestehen von Fehlern verminderte die Distanz zwischen Lernenden und Lehrenden in diesem Prozeß. Die Schüler erkannten unser Bemühen und gaben uns Feedback: „Wir wissen jetzt, daß Ihr uns mit Prüfungen nicht über den Tisch ziehen wollt.“ Dieser Vertrauensaufbau zum Thema Prüfung allgemein war das wichtigste Ergebnis für mich und gibt mir im weiteren Ausbildungsverlauf mehr Sicherheit im Umgang mit Leistungsbewertung.

Taxonomiestufen als Hilfe bei der Konstruktion von Aufgaben

Die Auseinandersetzung mit den Bloomschen Taxonomiestufen hilft mir bei der Konstruktion weiterer

Prüfungsaufgaben und gibt mir auch eine Möglichkeit, Aufgaben mit verschiedenem Schwierigkeitsgrad zu stellen. Z.B. findet demnächst eine mündliche Prüfung statt, und wir haben uns überlegt, die Schüler zwei Fragen über Los ziehen zu lassen, wobei die erste eine eher leichte Wissensfrage und die zweite eine anwendungsbezogene oder analytische Frage sein soll. So wird zum einen die Lernleistung belohnt, zum andern den Schülern die Möglichkeit gegeben, theoretisches Wissen auf konkrete Probleme zu übertragen, was einer höheren Leistung entspricht. Mit dieser Art der Prüfung kann ich mich sehr identifizieren, zumal sie dem Therapiealltag einer Logopädin zu entsprechen scheint, die immer wieder mit ihrem Wissen Lösungsansätze für das eigene Vorgehen kreieren muß.

Probleme als Lernprozess für alle

Daß auch mal was schief gehen kann bei solch hehren Projekten, zeigte die Durchführung im zweiten Kurs. Leider war hier der Leistungsdruck zunächst größer als die Offenheit gegenüber uns Lehrlogopäden. Zum anderen war die Information unsererseits nicht ausführlich genug,

so daß die Schüler nicht wußten, daß sie die gleiche Klausur wie Kurs 1 schreiben sollten. Wir hatten ja versucht, einige Fehler aus der ersten Version zu verbessern und freuten uns auf diesen neuen Durchgang. Da sich nun diese Klasse die Aufgabenstellung nebst Erwartungshorizont vom ersten Kurs besorgt hatte (mit der verständlichen Begründung, mal zu wissen, wie eine solche Klausur aussehen könnte), war natürlich die Eigenbewertung, sowie die Erforschung des veränderten Erwartungshorizontes und der besser gewichteten und offeneren Punkteverteilung sinnlos. Nachdem die erste Enttäuschung über das Scheitern verfliegen war, wurden jedoch auch im zweiten Kurs Diskussionen über Vertrauen und Offenheit geführt, die im nachhinein sehr gewinnbringend erscheinen und hoffen lassen, daß für die Zukunft eine Nähe von Studierenden und Lehrlogopäden im Bewertungsprozeß möglich wird. Tatsache ist, daß trotz vieler guter vertrauensschaffender Maßnahmen der Leistungsdruck hoch ist, und die Angst vor Bewertung und Prüfung in vielen sehr alt zu sein und tief zu sitzen scheint.

Ann Flügel, Lehrlogopädin

Die Stellungnahme von Magdalena Zunfmeister aus dem 1. Kurs paßt gut an das Ende dieses Erfahrungsberichtes. Sie nimmt eine übergeordnete Perspektive ein, die wie ein „Schlußwort“ die Sache abrundet, indem sie die relative Offenheit und den Prozesscharakter des Bewertungsproblems innerhalb unserer Logopädieausbildung anspricht:



Bewerten hat Bedeutung

Nach einem Jahr läßt sich sagen, daß sich das Thema als viel weitgreifender erweist, als es zunächst den Anschein hatte, und daß sich in dieser Sache kein abschließendes Resümee ziehen läßt, denn wir befinden uns noch mitten im Prozeß.

Warum? Bewerten heißt, wir messen jemanden oder etwas aufgrund unseres individuellen Bezugssystems (Fachwissen, Persönlichkeitsbezug usw.) Bedeutung zu. Dies geschieht aber nicht nur in Prüfungssituationen,

sondern jeden Augenblick, in dem wir einen Denkvorgang oder eine Handlung vollziehen. Diese Bedeutungsgebung hat also existentiellen Charakter, schlichtweg macht er unsere Individualität aus, und deswegen ist im besonderen das Thema Leistungsbewertung krisenverdächtig.

Das Klausurprojekt hatte zwar einen festumgrenzten Rahmen, in dem wir uns mit den Dozenten zum Thema Bewertung auseinandersetzten, jedoch wurde uns auch deutlich, daß diese Begrenzung eine Illusion ist, denn es war klar, daß

Korrelation Ko-Flü; Korrelkoeff = 0,9115

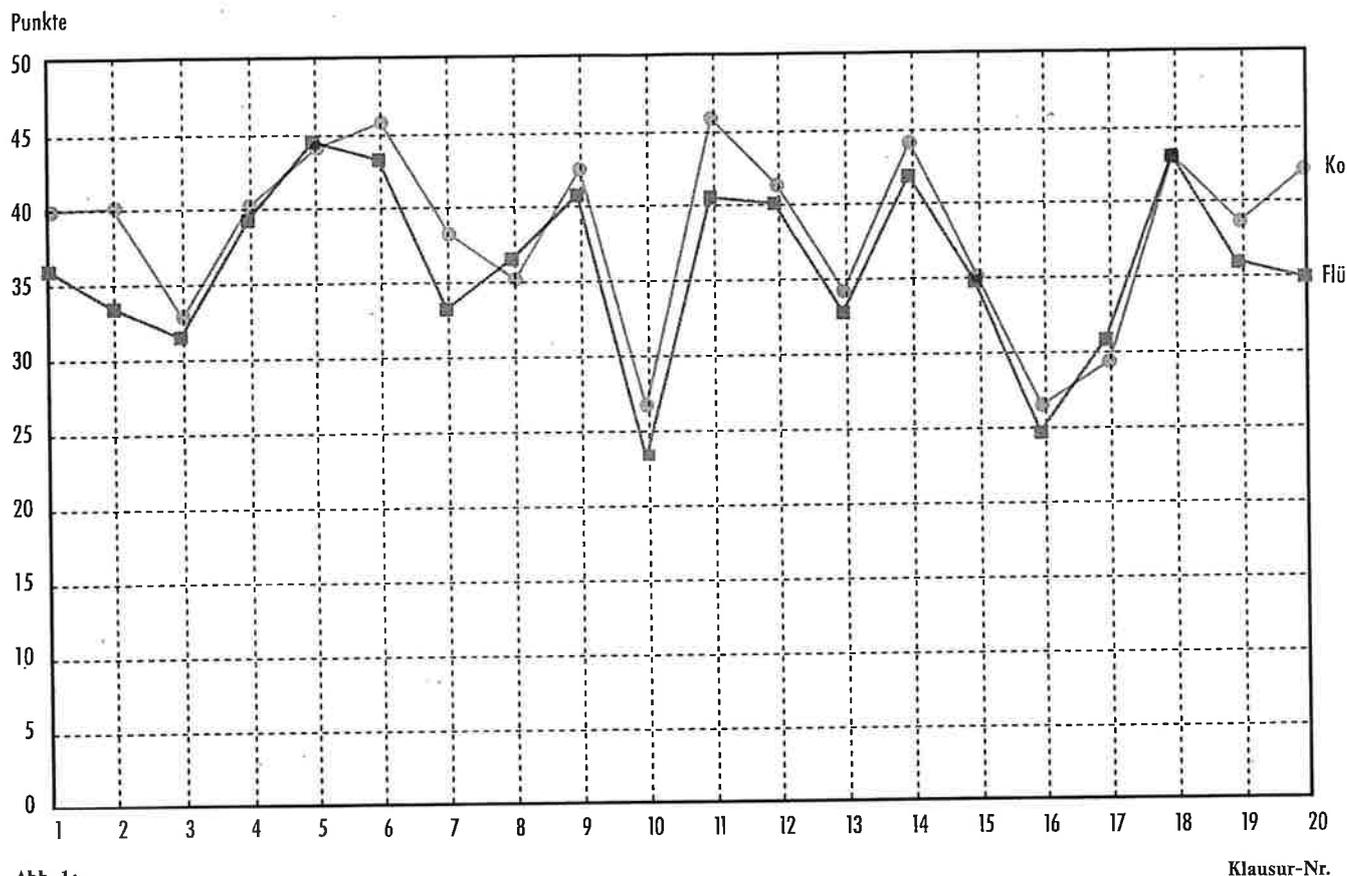


Abb. 1:

Ko = Bewerter I, Flü = Bewerter II

wir uns nicht nur auf eine Metaebene begeben und über Bewertung nachdenken, sondern daß alle Beteiligten bereit sein müssen, in einen vielschichtigen Umwertungsprozess einzusteigen. Das Projekt sprengte zum Teil den Rahmen der üblichen Schulhierarchie, plötzlich sollten wir Schüler den Notengebungprozess mitgestalten. Doch gleichzeitig können Themen wie Leistungsdruck, eingefleischte Prüfungsängste, Selbstwerteinschätzungen, Gruppenkonformität, Rollenerwartungen nicht im Schnellkurs bewältigt werden, bzw. das langsam wachsende Vertrauen, das durch die Dialoge mit den Dozenten erwuchs, steht immer wieder erneut auf dem Prüfstand aufgrund eben dieser äußeren

und inneren Vorbedingtheiten. So kann ich nur feststellen, daß das Projekt kein punktuellere Ereignis bleiben darf, daß zumindest alle Lehrlogopäden sich diesem Prozess anschließen sollten, um ein breiteres Spektrum auf schulischer Ebene zu erreichen. Zum zweiten brauchen wir Geduld, um Rückschläge einzustecken, die entstanden sind, weil Vertrauen einerseits und Ungläubigkeit gegenüber dem, was ich noch nicht erfahren habe, dialektische Kontrahenten sind. Schüler: „Kann es sein, daß der mich fragt, ob die Note, die er mir gemacht hat, mit der übereinstimmt, die ich mir selbst geben würde?“
- Ne, das glaube ich nicht, ... oder?“

Magdalena Zunftmeister, 1. Kurs

L I T E R A T U R

Glindemann, R./Springer, L. 1989: PACE-Therapie und sprachsystematische Übungen - Ein integrativer Vorschlag zur Aphasietherapie. In: Sprache - Stimme - Gehör 13. Stuttgart: Thieme-Verlag

Ingenkamp, K. (Hrsg.) 1971: Tests in der Schulpraxis. Weinheim: Beltz

Kotten, A. 1983: Aphasietherapie: Kommunikation oder Drill? In: Allhoff, Dieter-W. Hrsg.: Mündliche Kommunikation: Störungen und Therapie. Frankfurt/M.: Scriptor-Verlag

Weigl, I. 1979: Neuropsychologische und psycholinguistische Grundlagen eines Programms zur Rehabilitation aphasischer Störungen. In: Peuser, G. (Hrsg.): Studien zur Sprachtherapie. München: Wilhelm Fink Verlag

Zimbardo, P.G. 1983: Psychologie. Heidelberg: Springer-Verlag

Hier wurde nur die im „empirischen Teil“ genannte Literatur erwähnt. Inhaltlich gilt auch die Literaturliste des „theoretischen Teils“ aus L.O.G.O.S. Nr. 2, 5. Jahrgang

Autor: Jürgen Köhler in Zusammenarbeit mit Ann Flügel, IB-Schule für Logopädie, Schloßstr. 1, 78479 Reichenau