

Hartmut Rehlich, Friedrich-Schiller-Universität Jena

Denkprozessorientierte Leistungsmessung mit Computerunterstützung

Kurzvorstellung eines Forschungsprojekts zu einem neuen Testdesign

1. Ziele

Die alte Binsenweisheit, daß man nicht alles gleichzeitig haben kann, trifft auch auf Testungen zu. Es gibt Zielvorgaben, die miteinander im Widerspruch stehen und prinzipielle Probleme bei der Formulierung von Testaufgaben. Vorgestellt wird ein Testdesign, das - angelehnt an Strukturvorteile mündlicher Prüfungen - zur Integration von Zielvorstellungen beitragen soll und die Einflußnahme unterschiedlicher Sprachgewohnheiten auf die Interpretation und Bewertung von Testbearbeitungen reduzieren will.

Dabei stehen die folgenden Fragen im Vordergrund:

- a) Wie kann man durch computergestützte normierte Testungen differenziertere Informationen über Denkprozesse und kognitive Strukturen einzelner Probanden erhalten, als dies mit „papierbasierten Massentestungen“ möglich ist und damit zu einer *besseren Einschätzung von Schülerleistungen* gelangen?
- b) Kann durch computergestützte Simulation nachhakender Kommunikationsprozesse die Rolle sprachlicher verstehensrelevanter Störfaktoren bei der Bewertung von Schülerleistungen verkleinert und damit der *Evaluationsvorgang objektiviert* werden?
- c) Wie viel Korrektur- und Auswertungsaufwand *kann durch Automatisierung gespart* werden?

2. Zum Konzept

Jede Form von (Mathematik-) Leistungstests für größere und vor allem heterogene Populationen beinhaltet Kompromisse.

Validität: Einerseits soll ein Test möglichst differenzierte Rückschlüsse auf das Wissens- und Könnensprofil der einzelnen Probanden (Gruppen) geben. Im Rahmen der Bildungsziele des Mathematikunterrichts gehört dazu dann auch, daß komplexere Aufgaben, die individuelle Entscheidungen über Bearbeitungswege erfordern, integriert werden. Ein aus vielen Einzelitems zusammengesetzter Multiple-choice-Test ist hinsichtlich eines Mathematikbildes, wie es z. B. in der „Standarddiskussion“ zum Ausdruck

gebracht wird (aber natürlich nicht neu ist), nicht valide. Mit der Integration komplexerer Aufgaben handelt man sich jedoch ein erhebliches Problem ein, das hier kurz umrissen wird:

Während meiner Tätigkeit als Lehrer und Fachvertreter für Mathematik habe ich erfahren, daß es eigentlich unmöglich ist, valide schulübergreifende Vergleichsarbeiten zu konzipieren. Das Problem liegt weniger darin, daß evtl. verschiedene Inhalte unterrichtet wurden, sondern eher darin, daß es von Schule zu Schule und von Klasse zu Klasse andere „Mikrokulturen“ gibt. Zwischen dem Lehrer und seiner Klasse bildet sich im Laufe der Zeit ein subtiles Netz ausgehandelter, nicht verbalisierter Abmachungen und Gewohnheiten, das den Umgang und das gegenseitige Verstehen steuert. Die Bedeutung eines Satzes gibt es nicht, jegliche Satzbedeutung ist eine soziale Konstruktion. Z. B. kann die Aufforderung „Begründe diesen Zusammenhang!“ - je nach Umfeld - als Aufforderung, einen strengen Beweis zu führen, oder als Bitte, einige plausible Argumente zu nennen, verstanden werden.

Die Formulierung der eingeforderten Leistung wird immer auf dem Hintergrund vorangegangener Erfahrungen und unterschiedlicher Sprachgewohnheiten von den Probanden interpretiert. Eine im Sinne des Testkonstruktors defizitäre Bearbeitung kann darauf beruhen, daß der Proband die Aufgabe lediglich anders oder gar nicht verstanden hat. Ein solches Item testet dann letztlich vor allem, ob der Proband die in der Frage kodierte Erwartungshaltung des Aufgabenstellers in dessen Sinne dekodiert. Hier sehe ich auch erhebliche Probleme bei allen großen internationalen Vergleichsuntersuchungen mit ihren sehr heterogenen Probandengruppen. In mündlichen Prüfungen nutzen kompetente Prüfer zur Klärung solcher Voraussetzungen daher ihre Routine einer „Verständnisvergewisserung durch geeignete Rückfragen“.

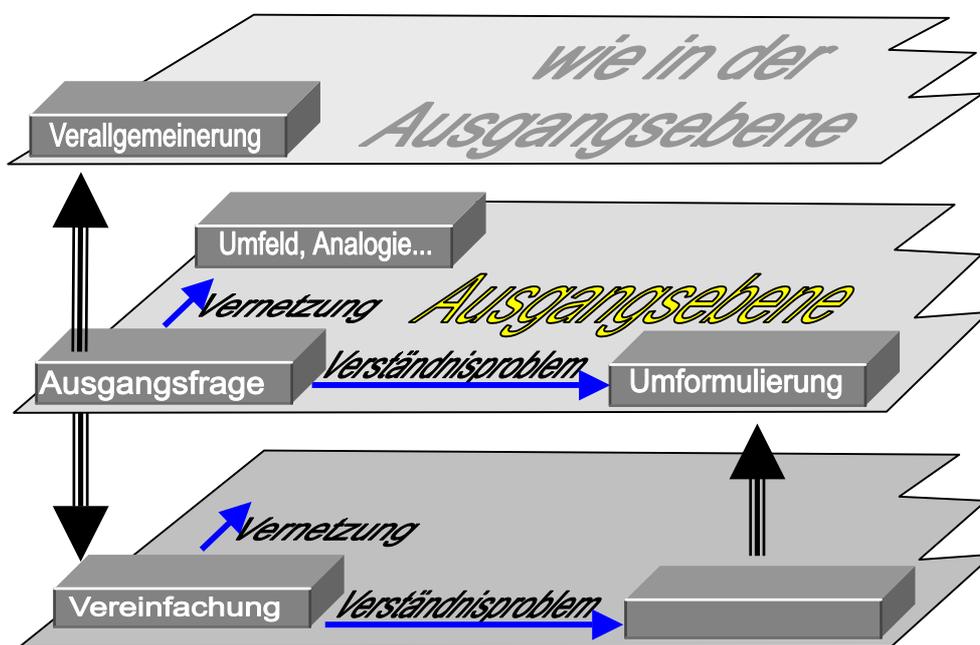
Auswertungsobjektivität: Andererseits soll sichergestellt sein, daß der Test „auswertungsobjektiv“ ist. Das ist schwer zu erreichen, wenn in das Testurteil auch die Würdigung individueller Bearbeitungswege komplexerer Aufgaben einbezogen werden soll und verschiedene Korrektoren mitarbeiten. Diese Zielvorgabe drängt daher eher zu „isolierten eindimensionalen“ Testitems oder rigiden Punkterastern für Einzelkomponenten vorgesehener mehrschrittiger Lösungswege. Diese Problematik kann auch nur bedingt durch ausführliche Evaluationsmanuale reduziert werden – abgesehen vom erheblichen Arbeitsaufwand für die Auswerter (siehe nächster Punkt).

Ökonomie: Die Gesamtheit der aufgewendeten Mittel muß in einem vernünftigen Verhältnis zum Ertrag stehen. Zu den Kosten gehört natürlich auch die Arbeitszeit verpflichteter Mitarbeiter (z. B. Lehrer), die evtl. zu-

sätzlich geleistet wird. Für die Qualität und Validität einer Untersuchung ist in der Regel die Akzeptanz bei allen Beteiligten nötig. Diese kann aber mit ökonomischen Bedingungen verknüpft sein.

3. Nachhakende Frageprozesse in einem Computerprogramm

Das skizzierte Dilemma kann etwas gemildert werden, wenn man „methodische Anleihen“ bei mündlichen Prüfungen macht. Neben der sehr grundsätzlichen Erwägung, daß mündliche Prüfer bei Problemen durch Rückfragen und daraus ggf. resultierende Umformulierungen der Fragestellung dafür sorgen können, daß sicher ist, daß der Proband die Frage verstanden hat, erhalten sie bei geeigneter Fragetechnik auch wesentlich differenziertere Informationen über das Wissens- und Könnensnetz des Geprüften. Durch das Ansprechen analoger Situationen kann die Übertragbarkeit mathematischen Wissens und Könnens auf andere mathematische Bereiche geprüft werden. Als „Einhilfen“ können Vereinfachungen vorgenommen werden und man kann auch prüfen, wie weit die Fähigkeiten des Probanden über die Ausgangsanforderung hinaus reichen. Die Grafik zeigt die Grundstruktur des entwickelten Computerprogramms.



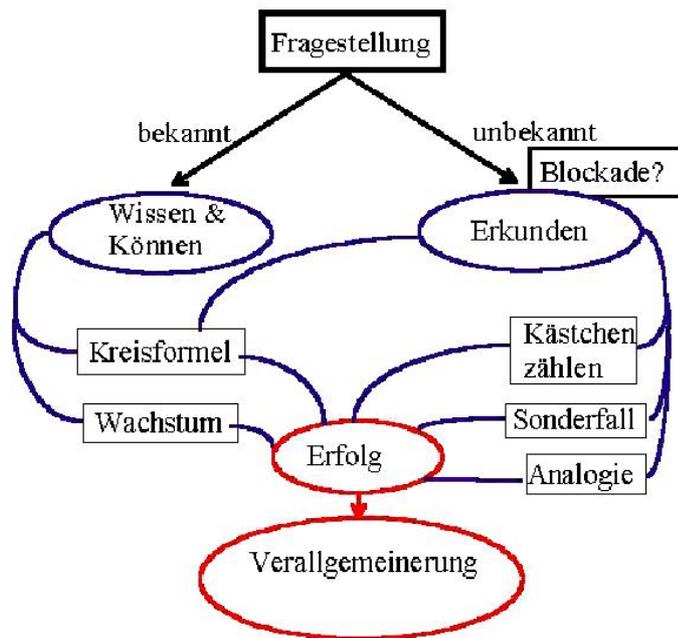
4. Ein Beispiel

Für die aus der PISA-Studie stammende „Pizzaaufgabe“ liegt exemplarisch eine ausgearbeitete Datenstruktur vor. Um die vielfältigen Probleme bei der ursprünglichen Aufgabe zu vermeiden, wurde eine etwas andere Ausgangsfrage gewählt:

Für den Basar an ihrer Schule stechen zwei Geschwisterkinder Kekse aus. Sie benutzen zwei kreisrunde Formen mit 4cm und 6cm Durchmesser. Von den größeren Keksen kommen jeweils vier in eine kleine Tüte. Wie viele der kleineren Kekse ergeben dasselbe Gewicht.

Das entwickelte Computerprogramm operiert auf der abgelegten Datenstruktur und erzeugt eine übersichtliche Oberfläche. Während der Bearbeitung werden den Probanden - abhängig von den individuell verschiedenen Eingaben - unterschiedliche Folgefragen vorgelegt.

Die Ausgangsfrage kann vom Probanden z.B. unreflektiert in das (oft funktionierende und – falls vorhanden - eigentlich positiv zu bewertende) „Metamuster Proportionalität“ eingeordnet werden. Dies ist ein ganz anderer Fehlertyp als z. B. ein schlichter Rechenfehler. Um weitere Aufschlüsse über die Kompetenz des Probanden zu erhalten, würde man in einer mündlichen Prüfung ganz unterschiedlich reagieren. Das wird im Programm nachgebildet. Im weiteren Verlauf kommt es zu zahlreichen weiteren Verzweigungsstellen. Das Schaubild zeigt eine erste grobe Differenzierung verschiedener Bearbeitungsverläufe. Zunächst muß geklärt werden, ob es sich für den Prüfling um eine Routineaufgabe handelt, da er vielleicht etwas Allgemeines über Wachstumsprozesse weiß, ob er eine Anwendungsaufgabe für Kreisformeln bearbeitet, oder ob er mit einer Problemlösung konfrontiert ist.



Allein das breite Spektrum unterschiedlicher Bearbeitungswege auf verschiedenem Niveau (bei einer im Sinne des Aufgabenkonstruktors verstandenen Aufgabe!) macht bereits deutlich, daß es nicht ausreicht, die Bearbeitungen von Aufgaben dieses Komplexitätsgrades nur nach Erfolg und Mißerfolg zu differenzieren.

Ausführlichere Informationen findet man in

Rehlich, Zimmermann: „Problem Solving in Mathematics Education. Proceedings of an International Symposium in September 2003“, Franzbecker 2004, ISBN 3-88-120-383-4.