

2.5 Zur Entwicklung kriteriumsorientiert normierter Leistungserhebungen auf der Grundlage eines Kompetenzebenenmodells

2.5.1 Problemlage

Als ein Merkmal von Leistungserhebungen wurde bereits bei allen betrachteten Funktionen im Abschnitt 2.2 die Art der Normierung diskutiert. Dabei zeigte sich, dass bis auf Erhebungen für die Ermittlung von Schülerleistungen, für die Ermittlung von Vergleichsdaten zur Fundierung bildungspolitischer Entscheidungen bzw. für regionale Schulaufsichtsbehörden eine kriteriumsorientierte Normierung notwendig oder zumindest sehr wünschenswert ist.

In der psychologischen Diagnostik ist die sozialnormorientierte Vorgehensweise zu Festlegen von Normen das übliche Vorgehen. Nach unserer Kenntnis basieren fast alle standardisierten psychodiagnostischen Verfahren, die gegenwärtig verfügbar sind, auf dieser Methode. Damit verbunden ist die grundlegende Herangehensweise der psychologischen Testtheorie, das *Verhalten von Personen* in bestimmten *Situationen* zu beobachten (Rost 2004, S. 24). Die aufgestellten Modelle und Normierungen beziehen sich deshalb immer auf die Verhältnisse in einer bestimmten Population und auf die im Test verwendeten Situationen, also die verwendeten Items.

Daraus ergibt sich u. a., dass es sehr unterschiedliche Modelle für einzelne Personenmerkmale gibt. Ein Beispiel dafür ist das räumliche Vorstellungsvermögen, für das verschiedene Erklärungsmodelle je nach verwendeter Testbatterie und getesteter Population existieren (Maier 1999).

Auch Bender (2005) weist daraufhin, dass das von der deutschen PISA-Gruppe auf der Grundlage der Daten des PISA-Testes 2000 entwickelte Kompetenzstufenmodell eigentlich nur für die Aufgaben dieses Testes gilt und sich bei einer Änderung der Leistungsvermögens in der betrachteten Schülerpopulation sich auch eine Änderung der Einstufung der Aufgaben ergeben könnte (S. 276).

Lehmann u. a. (2001) betonen ausdrücklich, dass mit den im Projekt LAU gebildeten Fähigkeitsniveaus nicht der Anspruch erhoben wird, eine mathematikdidaktische oder entwicklungspsychologisch begründete Theorie der Lernfortschritte zu präsentieren. Es handele sich lediglich um psychometrisch abgegrenzte Bereiche der Metrik, auf die sich die Fähigkeits- und Schwierigkeitsparameter beziehen. Die inhaltliche Interpretation sei allein Aufgabe der Mathematikdidaktik und nicht der empirischen Bildungsforschung (S. 35).

Bei der Anwendung der sozialnormorientierten, d. h. am aktuellen Verhalten der Schüler ausgerichteten Vorgehensweise der Psychodiagnostik auf die Messung von Schulleistungen im Fach Mathematik ergeben sich zumindest folgende grundsätzliche Probleme.

Das Verhalten eines Schülers beim Lösen einer mathematischen Aufgabe ist in der Regel Resultat einer großen Vielzahl von einzelnen kognitiven Dispositionen. Dies wird bereits durch sachlogische, epistemologische oder linguistische Analysen der jeweiligen Anforderungen deutlich. In diesem Buch erfolgt eine solche Analyse am Beispiel einer Aufgabe im Kapitel 3. Meyerhöfer (2004 b) hat diesen Sachverhalt ebenfalls in einer umfangreichen Analyse möglicher Lösungswege von PISA-Aufgaben untersucht.

Ein weiteres Problem ergibt sich aus der Tatsache, dass im Unterschied zu den meisten sonstigen Anwendungen der Psychodiagnostik der Leistungsstand eines Schülers nur einen momentanen zeitlichen Zustand beschreibt, der in erheblichem Maße von dem vorherigen Unterricht abhängt und sich im weiteren Unterricht sehr bald wieder ändert. Es ist deshalb unmöglich, den Test unter den gleichen Bedingungen zu wiederholen.

Mit einer Leistungsdiagnostik im schulischen Bereich ist weiterhin immer die Frage verbunden, ob und inwiefern der ermittelte Leistungsstand den curricularen Erwartungen entspricht. Dies erfordert eine lehrziel- also kriteriumsorientierte Normierung.

Mit diesen Problemen ist eine aktuelle Hauptfrage der Kompetenzdiagnostik verbunden, theoretisch fundierte Kompetenzmodelle als Grundlage für psychometrische Messverfahren zu entwickeln. Kaiser (2000) stellt heraus, dass eine inhaltliche Theorie nötig ist, die das Verhältnis zwischen dem bei TIMSS und PISA als Konstrukt verwendeten latenten Merkmal "mathematical literacy" und den beobachtbaren Indikatoren erklärt. Eine solche Theorie existiert nach ihren Worten derzeit noch nicht, ebenso wenig wie ein Konsens innerhalb der Mathematikdidaktik über das theoretische Konstrukt. Sie fordert, dass durch die Mathematikdidaktik als inhaltliche Disziplin solche Basistheorien vor der Entwicklung von Test bereitzustellen sind, um zu erreichen, dass die Items nach inhaltlichen Anforderungen ausgewählt werden. Die aktuellen kontroversen Diskussionen zu den so genannten Kompetenzstufen, Kompetenzniveaus bzw. Kompetenzklassen (Bender 2005, Meyerhöfer 2004, Lind u. a. 2005) spiegeln die unzureichenden theoretischen Grundlagen wider.

In ihrem Antrag zur Einrichtung eines Schwerpunktprogramms zur Entwicklung von Kompetenzmodellen Bezeichnen Klieme und Leutner (2006) die Entwicklung theoretisch begründeter und empirisch geprüfter Kompe-

tenzmodelle als bisher noch nicht zufrieden stellend gelöstes Problem. Bei der Entwicklung konkreter Aufgabenstellungen zur Diagnostik spezifischer fachbezogener Kompetenzen würde häufig deutlich, dass der Abstraktionsgrad der bisher entwickelten Kompetenzmodelle zu hoch sei. Den Testentwicklern bleibt dann keine andere Wahl, als sehr große Mengen von Aufgaben zu entwickeln, diese Aufgaben empirisch zu erproben und dann im Nachhinein diejenigen Aufgaben beizubehalten, die dem jeweils verwendeten Testmodell entsprechen. Auf diese Weise wären z. B. die bei PISA international berichteten Kompetenzstufenmodelle entstanden. Die Kompetenzstufen wurden dabei allerdings nicht a priori theoretisch begründet, sondern nach Sichtung der testmodell-konform übrig gebliebenen Aufgaben post-hoc definiert, was aus theoretischer Perspektive wenig befriedigend sei. (S. 7)

In den 1970er Jahren wurden zahlreiche Konzepte der kriteriumsorientierten Leistungsmessung entwickelt (Klauer 1987) und im Rahmen der damaligen lehr- und lernzielorientierten Curriculumsreformen eingesetzt. Diese sind letztlich gescheitert und damit wurden auch die Messkonzepte nicht weiter entwickelt. Eine Ursache für die durch die pädagogisch-psychologische Diagnostik mit verursachte Fehlsteuerung von Bildungsprozessen sehen Klieme und Leutner (2006, S. 5) in der fehlenden Differenzierung nach Domänen und Erhebungszwecken und der fehlenden Verbindung von kognitiver Modellierung und Psychometrie.

Die Entwicklung kriteriumsorientierter Testverfahren ist ein sehr komplexes und aufwendiges Vorhaben. Die große Anzahl von Freiheitsgraden, die man bei normativen Festlegungen hat, darf nicht darüber hinwegtäuschen, dass es wesentliche Restriktionen gibt, die man als „pädagogische Erhaltungssätze“ bezeichnen könnte: Man muss davon ausgehen, dass die Summe der aufgewendeten „Arbeit“ (als geistig-praktische Tätigkeiten) von Schülern und Lehrern sowie die Summe der aufgewendeten Arbeitszeit (Lerngelegenheiten) im Wesentlichen stets konstant sind. Weiterhin ist davon auszugehen, dass für einen bestimmten historischen Zeitraum auch die Summe aller möglichen Einzelleistungen von Schülern bzw. Lehrern (als Quotient von aufgewendeter Arbeit pro Zeit) bedingt durch das vorhandene allgemeine Leistungsvermögen von Schülern und das pädagogische Können von Lehren relativ konstant ist.

Deshalb ist bereits die theoretische Optimierung der Summe aller Einzelleistungen ein hochkomplexes Problem. Sie ergibt sich praktisch aus den vorherrschenden Gestaltungsweisen des Unterrichts und dem damit verbundenen aktuellem pädagogischen Können der Lehrer sowie weiteren Parametern, die sich nur über einen größeren Zeitraum ändern lassen. Wenn

auf irgendeinem Gebiet ein höherer Aufwand betrieben wird, verbessern sich naturgemäß die Leistungen der Schüler auf diesem Gebiet. Auf Grund der genannten „Erhaltungssätze“ von Arbeitsumfang, Zeitaufwand und Leistungsfähigkeit bedeutet dies aber meist automatisch eine Verschlechterung der Leistungen auf anderen Gebieten.

2.5.2 Zur Entstehung und den Quellen des Kompetenzebenenmodells

Wie bereits dargestellt (s. S. 40 ff.) hatten die Leistungserhebungen in Mecklenburg-Vorpommern die bildungspolitische Zielstellung, auf die massiven Klagen aus der beruflichen Ausbildung über mangelnde mathematische Grundkenntnisse und Grundfertigkeiten zu reagieren. Dies hatte zur Folge, dass das Anforderungsniveau der Aufgaben bewusst niedrig gehalten wurde. Von ministerieller Seite wurde auf „unteres Hauptschulniveau“ orientiert. Die dadurch vorgenommene Beschränkung der Anforderungsbreite des mathematischen Wissens und Könnens der Schüler wurde zunächst als "grundlegendes Wissen und Können" bezeichnet. Im Zuge der weiteren Arbeiten entstand im ständigen Wechselverhältnis von theoretischen Überlegungen und Auswertung der empirischen Ergebnisse ein Konzept mathematische Bildung, das wir dann später um Verwechslungen vorzubeugen als „sicheres Wissen und Können (SWK)“ bezeichnet haben.

Bei unseren Überlegungen berücksichtigten wir zahlreiche ähnlich gelagerte Bestrebungen. So gab es bei der Entwicklung der Bildungsstandards Diskussionen über das in den Standards fest zu schreibende Niveau. Die Verfasser der Expertise zur Entwicklung nationaler Bildungsstandards (Klieme u. a. 2003) plädierten nachdrücklich für die Festlegung von Mindeststandards, die sie für die Qualitätssicherung im Bildungswesen von entscheidender Bedeutung halten (S. 20) während von der KMK entschieden wurde, Regelstandards zu verfassen.

Auf die Notwendigkeit einer Ausweisung von Mindestkompetenzen weisen auch Stern, Hardy (2001) hin. Sie betonen, dass eine vergleichende Leistungsmessung sich nicht nur auf die Ermittlung von Durchschnittsleistungen beschränken sondern auch Aussagen darüber beinhalten sollte, welcher Anteil der Schüler über eine Mindestkompetenz verfügt. Die Festlegung von Kriterien für Mindestkompetenzen stellt für sie eine zentrale Herausforderung aller an der Konstruktion von Leistungstests beteiligten Personen dar und muss zwischen Lehrpersonen, Bildungspolitikern und Wissenschaftlern diskutiert werden (S. 158).

Auch in anderen Bundesländern gibt es Bestrebungen, im Rahmen von Leistungserhebungen einen Schwerpunkt auf die Sicherung von Mindest-

standards zu legen. So soll mit den Erprobungsarbeiten an Mittelschulen in Sachsen den „... immer wieder von der Wirtschaft erhobenen Forderungen nach einem größeren Stellenwert grundlegender Kompetenzen wie Kopfrechnen, Fähigkeiten im schriftlichen Sprachgebrauch und Kommunikationsfähigkeit nachgegangen“ werden (<http://www.sachsen-macht-schule.de/erprobungsarbeiten/index.html>)

Im Lehrplanentwurf für das achtjährige Gymnasium in Bayern werden in den Vorbemerkungen ausführliche Betrachtungen zu einem „Grundwissen“ angestellt. Dabei versteht man unter Grundwissen „nicht nur Wissen, sondern auch Fähigkeiten, Fertigkeiten oder sogar Haltungen“ und betont, dass man diesen Begriff nur aus Rücksicht auf seine starke Verwurzelung in der Bildungsdiskussion verwendet, ansonsten würde man ihn durch den Begriff „Kern-Kompetenzen“ ersetzen. In allen Jahrgangsstufenplänen der einzelnen Fächer sind nach dem einleitenden Zieltext die Lernziele und Lerninhalte angegeben, die als Grundwissen nachhaltig gelernt und beherrscht werden müssen. Das so bestimmte Grundwissen soll auch noch nach Jahren, d. h. auch nach der Schulzeit parat sein.

(<http://server.co101.spacenet.de/isb.co101.spacenet.de/contentserv/3.1/g8.de/index.php?StoryID=1>)

In zahlreichen Fachschaften von bayrischen Gymnasien wurden Grundwissenskataloge erarbeitet, die im Internet angeboten werden. In einem Schulversuch von 1997 - 2002 haben 7 Gymnasien in arbeitsteiliger Weise für einzelne Stoffgebiete dieses Grundwissen formuliert und durch Aufgaben konkretisiert. Als Kriterien zur Auswahl des "Grundwissens" verständigten sich die Fachschaftsvertreter auf folgende drei Felder:

- Anforderungen der folgenden Jahrgangsstufen (auch in anderen Fächern)
- Anforderungen im Studium, Beruf, Alltag
- Bildungswert der Mathematik an sich (Problemanalyse, Entwicklung von Problemlösestrategien, Argumentationsfähigkeit, Logik, kulturgeschichtliche Bedeutung ...)

(<http://www.isb.bayern.de/gym/allgem/math/grundwissen.htm>)

Auch für die bayrischen Realschulen wurden im Rahmen von Projekten Materialien zum Grundwissen für einzelne Klassenstufen erarbeitet. (<http://www.isb.bayern.de/isb/index.asp?MNav=5&QNav=6&TNav=0&INav=0&Fach=30>)

Unser Konzept des sicheren Wissens und Könnens haben wir nach Abschluss der unmittelbaren Auswertungen der Vergleichsarbeiten in gemeinsamer Arbeit mit Gruppen interessierter Lehrer auf drei Bereiche mathema-

tischer Bildung angewendet: das Wissen und Können im Arbeiten mit Größen, das Wissen und Können in der ebenen Geometrie sowie in der räumlichen Geometrie. Ergebnisse der intensiven Zusammenarbeit, die sich jeweils über ein Schuljahr erstreckte, waren Standpunkte zum Umfang und Niveau der betreffenden Anforderungen, die auch in Form von Aufgaben konkretisiert wurden. Die dabei entstandenen Broschüren wurden mit Unterstützung des Landesinstitutes für Schule und Ausbildung Mecklenburg-Vorpommern an alle Schulen des Landes verteilt und werden gegenwärtig erprobt (Sill u. a. 2004, Sill u. a. 2005 a, Sill u. a. 2005 b). In Anpassung an die neuen Sprachregelungen haben wir das Konzept des sicheren Wissens und Könnens bei diesen Arbeiten auch als Mindeststandards bezeichnet, obwohl es sich zum Teil davon unterscheidet. (Sill 2006, S. 315 ff.)

Um diese Unterschiede, insbesondere zur Stufung nach Anforderungsniveaus deutlich zu machen, entstand in jüngster Zeit die Idee, verschiedene *Kompetenzebenen* zu unterscheiden. Befördert wurde diese Idee durch unsere aktuellen Bemühungen, zusammen mit erfahrenen Lehren die allgemeinen Zielstellungen des neuen Kerncurriculums für die Qualifikationsphase der gymnasialen Oberstufe zu präzisieren. Durch die vor allem politisch motivierte Erstellung eines gemeinsamen Planes für drei Bundesländer (Berlin, Brandenburg und Mecklenburg-Vorpommern) trotz völlig unterschiedlicher Rahmenbedingungen ist in Mecklenburg-Vorpommern die Situation entstanden, bei einem 12jährigen Abitur ohne Differenzierung in Grund- und Leistungskurse den gleichen Rahmenplan für je 4 Wochenstunden in der Qualifikationsphase zu haben wie ein Land mit 13jährigen Abitur für den Leistungskurs mit je 5 Wochenstunden.

2.5.3 Grundlagen, Merkmale und Funktionen eines Kompetenzebenenmodells

Der aktuelle Stand unserer konzeptionellen Überlegungen lässt sich etwa in folgender Weise beschreiben, wobei aus Umfangsgründen unsere theoretischen Ausgangspositionen nur knapp dargestellt.

Theoretische Grundlagen des Kompetenzebenenmodells

Wir unterscheiden zwischen dem **Anforderungsniveau** und dem **Schwierigkeitsgrad** einer Aufgabe. Beide Begriffe werden oft nicht klar voneinander abgegrenzt und teilweise als Synonyme verwendet. So sprechen Helmke; Hosenfeld (2004) bei ihren konzeptionellen Überlegungen zum Projekt VERA von schwierigkeitsbestimmenden Merkmalen und meinen damit Merkmale wie die Textlänge und die Anzahl gegebener Informationen, die das Anforderungsniveau von Aufgaben charakterisieren. Die Validität dieser Merkmale überprüfen sie anhand der Daten aus Stichproben.

Unter dem *Anforderungsniveau* einer Aufgabe verstehen wir ein qualitatives Merkmal, das durch die Anzahl und die Komplexität der zum Lösen der Aufgabe notwendigen geistigen und geistig praktischen Handlungen charakterisiert wird. Damit eng verbunden sind Analysen zu den möglichen Lösungswegen und den möglichen fehlerhaften Gedankengängen beim Bearbeiten der Aufgabe. Die Ermittlung des Anforderungsniveaus erfordert u. a. sachlogische und linguistische Analysen der Aufgabenstellung. Wir geben in diesem Buch im Kapitel 3 eine mögliche generelle Schrittfolge zu Ermittlung der Gesamtheit der Anforderungen und der dabei möglichen Fehler an. Die allgemeine Beschreibung von Merkmalen bestimmter Aufgabentypen zur Klassifizierung von Niveaustufen der Anforderung ist allerdings ein sehr schwieriges und noch weitgehend ungelöstes Problem. Bei der Bestimmung des Anforderungsniveaus von Aufgaben im Mathematikunterricht bleibt der mögliche Bearbeiter der Aufgabe weitgehend außerhalb der Betrachtungen. Man hat sich einen Schüler vorzustellen, der über alle kognitiven, affektiven und motorischen Voraussetzungen zum Lösen der Aufgabe verfügt. Da sich die meisten Aufgaben einem bestimmten Entwicklungsstand der mathematischen Leistungsfähigkeit von Schülern zuordnen lassen, basieren die Überlegungen zum Anforderungsniveau oft indirekt auf der Annahme einer bestimmten Schülerpopulation als Löser der Aufgabe.

Unter dem *Schwierigkeitsgrad* einer Aufgabe verstehen wir die empirisch ermittelte Erfüllungsquote bzw. daraus abgeleiteter Parameter (Schwierigkeitsindex, Aufgabenindex) beim Lösen dieser Aufgabe in einer bestimmten Schülerpopulation. Der Schwierigkeitsgrad ist also eine subjektive und temporäre Kategorie, deren Angabe nur Sinn macht im Zusammenhang mit der genauen Beschreibung der Bedingungen des voran gegangenen Unterrichts, der Angabe der getesteten Aufgaben, des Zeitpunktes der Erhebung und der getesteten Population.

Betrachtet man eine bestimmte Schülerpopulation zu einem bestimmten Zeitpunkt, so sind das Anforderungsniveau und der Schwierigkeitsgrad einer Aufgabe in der Regel voneinander abhängig, d. h. je höher das Anforderungsniveau einer Aufgabe ist umso geringer wird ihre Lösungswahrscheinlichkeit in der Gesamtpopulation und umso höher ist auch ihr Schwierigkeitsgrad in dieser Population. Betrachtet man aber den Schwierigkeitsgrad bezogen auf Teilmengen der Population bis hin zu einzelnen Schülern, so ändert sich der Schwierigkeitsgrad bei gleichem Anforderungsniveau in Abhängigkeit vom vorhergehenden Unterricht. Dieselbe Aufgabe kann für einzelne Schüler oder Schülergruppen einer Klasse von völlig unterschiedlicher Schwierigkeit sein.

Damit im Zusammenhang steht auch eine andere Sicht auf die Befähigung zum Problemlösen. Unter einem Problem sollte nach unserer Auffassung eine Aufgabe verstanden werden, die für einen Schüler nicht unmittelbar lösbar ist. Mit dieser Begriffsfassung ist der Terminus Problem eine subjektbezogene Kategorie d. h. nur im Zusammenhang mit der Betrachtung der Person, die die Aufgabe lösen soll, sinnvoll. Auch eine sehr einfache Aufgabe kann ein Problem für einen Schüler aufgrund noch nicht vorhandener oder schwach ausgebildeter Leistungsdispositionen sein. Die Befähigung zum Problemlösen umfasst damit nicht nur die Behandlung besonders anspruchsvoller Aufgaben.

Betrachtet man die Leistungsfähigkeit einer Schülerpopulation in ihrer zeitlichen Entwicklung, so können sogar das Anforderungsniveau und der Schwierigkeitsgrad bezogen auf die Gesamtpopulation völlig auseinander fallen. Es ist durchaus möglich, den Schwierigkeitsgrad einer Aufgabe bedeutend zu senken, ohne ihr Anforderungsniveau zu verringern, indem dieser Aufgabentyp ausgiebig und intensiv im Unterricht behandelt wird. Hier werden die Grenzen einer sozialnormorientierten Kompetenzstufung deutlich, bei der sich die Kompetenzstufen aus der aktuellen Leistungsfähigkeit der Population ergeben.

Eine weitere Grundlage ist die Betrachtung von **Qualitätsparametern psychischer Disposition**. So können zum Beispiel Elemente des Wissens gekennzeichnet werden durch den Grad ihrer Allgemeinheit, Systemhaftigkeit, Anschaulichkeit, Dauerhaftigkeit, Sinnhaftigkeit oder Disponibilität (Pippig 1985). Die Ausprägung der Qualitätsparameter kann mit geeigneten Methoden überprüft werden.

Wird z. B. der Begriff Quader betrachtet, so zeigt sich der Grad der Allgemeinheit u. a. darin, ob ein Schüler auch einen Würfel als Quader bezeichnet, ob er Kantenmodelle, Vollkörper oder Flächenmodelle richtig identifiziert und den Begriff auch für quaderförmige Flüssigkeits- oder Gasmenngen anwenden kann. Eine Grundlage für die Untersuchung der Systemhaftigkeit des Quaderbegriffs ist die Betrachtung des semantischen Netzes (Klix 1984), das den Begriff Quader als einen Knoten enthält.

Wir verstehen psychische Dispositionen in ihrer Repräsentation im Gehirn als räumlich-zeitliche Aktivitätsmuster neuronaler Netze, die durch Langzeitpotenzierung entstehen. Aus dieser Sichtweise halten wir die Verfestigung bzw. Verfügbarkeit der psychischen Dispositionen als einen für Unterrichtsprozesse wichtigen Parameter. So gibt es z. B. Gedächtnisinhalte eines Schülers, über die er jederzeit verfügen kann und auch solche, die ihm zunächst nicht bewusst sind, aber durch eine Wiederholung bzw. Re-

aktivierung wieder in Erinnerung gerufen werden können. Wie schon bei den Betrachtungen zum Verhältnis von Anforderungsniveau und Schwierigkeitsgrad diskutiert, wollen wir auch die zeitliche Dimension in das Kompetenzmodelle einbeziehen und somit die Prozesse des Vergessens und Reaktivierens berücksichtigen.

Ein weiterer Ansatz unserer Betrachtungen ist eine **Gewichtung der Bildungsziele** des betreffenden Unterrichtsfaches, als entscheidender Schritt zu einer kriteriumsorientierte Normierung. Wir fassen Bildungsziele als intendierte psychischer Dispositionen bzw. Systeme von Dispositionen auf. Unter Gewichtung verstehen wir die Ausweisung von anzustrebenden Qualitätsparametern der betreffenden Dispositionen. Eine solche Gewichtung eines Ziels ergibt sich aus seiner Bedeutung in folgenden Bereichen:

- (1) der weiteren Unterricht in diesem und in anderen Fächern,
- (2) die weitere schulische bzw. berufliche Ausbildung,
- (3) die konkrete Anforderungen im Alltag von Bürgern,
- (4) die Anforderungen an die Allgemeinbildung eines Bürgers.

Dies entspricht z. B. den Betrachtungen im oben erwähnten Projekt bayerischer Gymnasiallehrer zur Bestimmung des Grundwissens (s. S. 127).

Bis auf den Aspekt der innerschulischen Bedeutung führen die Betrachtungen zu den übrigen drei Aspekten über den Rahmen der Schule hinaus. Es sind Untersuchungen in zahlreichen gesellschaftlichen Bereichen erforderlich, wie sie auch Heymann (1996, S. 135 ff.) zitiert bzw. selbst durchgeführt hat. Die Anzahl der von Heymann angegebenen Untersuchungen ist aber sehr gering und die Ergebnisse sind offensichtlich noch wenig konkret. Bei unseren Projekten zur Gewichtung der Ziele bezüglich des Wissens und Könnens im Arbeiten mit Größen sowie des geometrischen Wissens und Könnens (vgl. S. 128) haben wir selbst lediglich zusammen mit Lehren eigene Überlegungen zur Bedeutung der Ziele in den vier genannten Bereichen angestellt und sind dabei von unseren Vorstellungen zu den Anforderungen in den angegebenen 4 Bereichen ausgegangen. Erforderlich sind also noch zahlreiche empirische Untersuchungen zur Struktur und dem Niveau der tatsächlichen Anforderungen.

Merkmale des Kompetenzebenenmodells

Die Grundidee besteht in der Konstruktion von Kompetenzebenen psychischer Leistungsdispositionen, die einen bestimmten Grad der Ausprägung von Qualitätsparametern der Dispositionen, insbesondere den Grad und die Art ihrer Verfügbarkeit erfassen sollen. Der Ausgangspunkt für die Bildung der Kompetenzebenen ist folgende Modellvorstellung zur realen Struktur

der psychischen Dispositionen in einem bestimmten Kompetenzbereich, wie etwa dem Wissen und Können im Arbeiten mit Größen (s. Abb. 2.1).

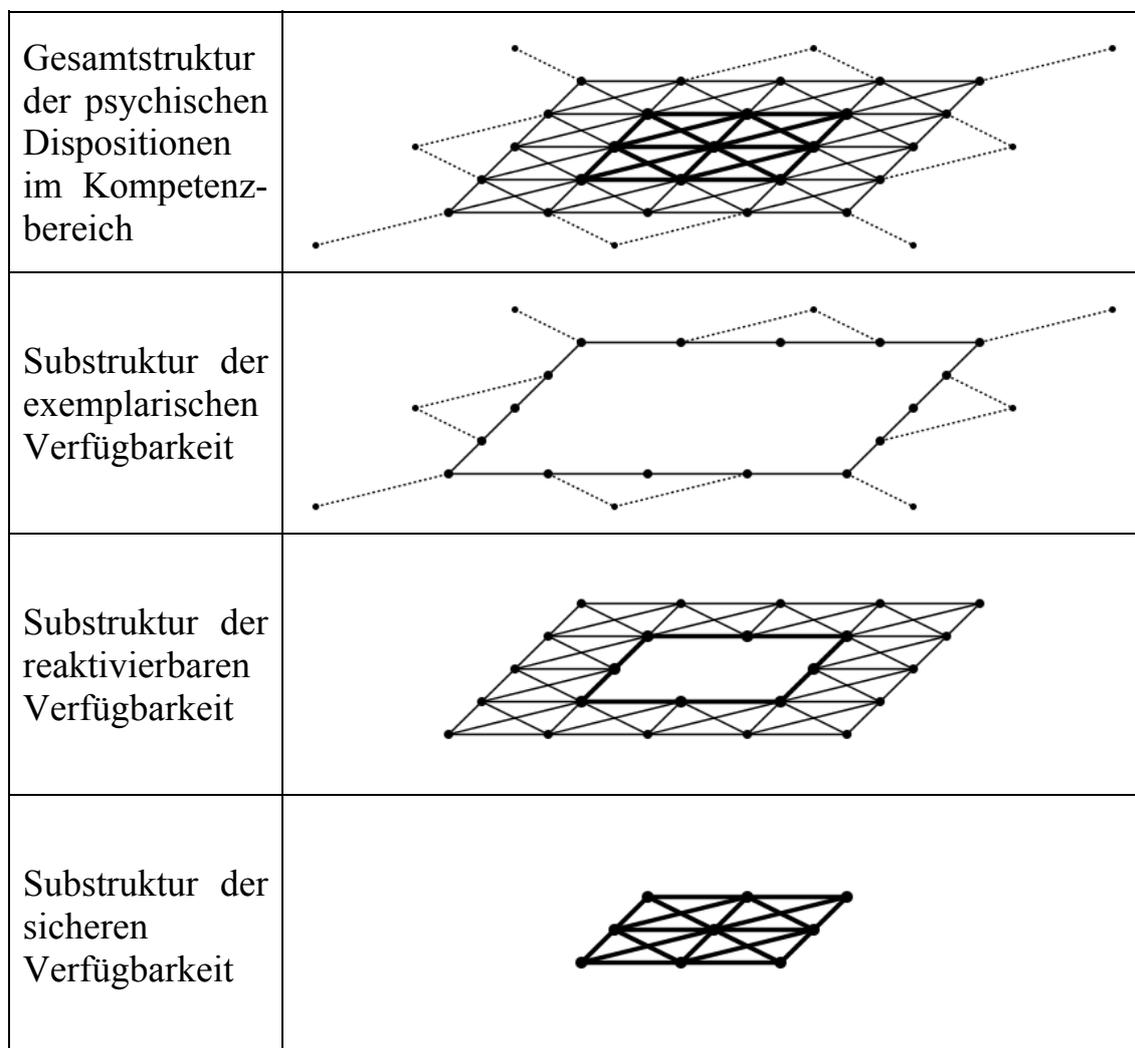


Abb. 2.1 Strukturen psychischer Dispositionen in einem Kompetenzbereich

Davon ausgehend unterscheiden wir folgende drei Kompetenzebenen der Ausprägung von Leistungsdispositionen.

- *Ebene der sicheren Verfügbarkeit:* Die Schülerinnen und Schüler verfügen über die Dispositionen jederzeit ohne eine spezielle Reaktivierung. Die Dispositionen werden als *sicheres Wissen und Können* bezeichnet.
- *Ebene der reaktivierbaren Verfügbarkeit:* Die Schülerinnen und Schüler verfügen über die Dispositionen am Ende eines Stoffgebiets. Nach einem gewissen Zeitraum ist jedoch eine Wiederholung und Reaktivierung des Wissens und Könnens erforderlich, bei der der schon einmal vorhandene Beherrschungsgrad wieder erreichbar ist.

- *Ebene der exemplarischen Verfügbarkeit:* Die Schülerinnen und Schüler haben erste Einsichten, Vorstellungen bzw. Fähigkeiten hinsichtlich des betreffenden Inhalts des Unterrichts. Sie können z. B. einfache Beispiele angeben, einige wichtige Merkmale oder Gedanken zu einer Vorgehensweise äußern und sich an entsprechende Episoden erinnern.

Auf allen Ebenen soll ein Schüler Aufgaben mit unterschiedlichem Anforderungsniveau und mit einer bestimmten Lösungswahrscheinlichkeit bearbeiten können, wobei der Anteil der anspruchsvollen Aufgaben von der Ebene der sicheren Verfügbarkeit zur Ebene der exemplarischen Verfügbarkeit zunimmt und die normativ festzusetzende Lösungswahrscheinlichkeit abnimmt. Die kognitive Struktur der Leistungsdispositionen, die sich auf der Ebene der sicheren Verfügbarkeit befinden, sollen den Schüler also in die Lage versetzen, auch bestimmte anspruchsvolle Aufgaben zu lösen, deren Beherrschung mit diesem Grad erforderlich ist. Damit unterscheidet sich unser Ansatz von den Bestrebungen, mit einem „Grundwissen“ bzw. einer Kompetenzstufe I nur sehr einfache Anforderungen zu erfassen. Mit den Kompetenzebenen wird also neben dem sozialnormorientierten Anspruchsniveau eine weitere Dimension eröffnet.

Das Ebenenmodell bezieht sich auf eine reale hirnspsychologische Konstellation und ist damit kein positivistisches Konstrukt für die Modellierung des Verhaltens von Personen. Dies impliziert die Notwendigkeit von empirischen Untersuchungen zum Wechselverhältnis von realer Struktur und Verhalten mit den Methoden verschiedener Wissenschaften.

Die Ebenen sollen neben der Struktur auch eine zeitliche Dynamik der psychischen Dispositionen eines Schülers erfassen, ohne dass dies in der schematischen Darstellung in Abb. 2.1 verdeutlicht werden konnte. Mit der Abbildung soll die Situation in einem Kompetenzbereich am Ende der Schulzeit nach Abschluss der Prozesse der Verarbeitung und des Vergessens veranschaulicht werden. Im Laufe der Herausbildung der Struktur gibt es bestimmte Zwischenstadien. So sollte etwa direkt nach einem Unterrichtsprozess zum Kompetenzbereich auch das Wissen und Können auf der Ebene der reaktivierbaren Verfügbarkeit sicher beherrscht werden.

Das Modell soll ebenfalls unterschiedliche Klassen von situationsspezifischen Anforderungen beschreiben, die mit unterschiedlichen Lernprozessen im Mathematikunterricht verbunden sind. So entspricht die Ebene der sicheren Verfügbarkeit den Situationen, in denen der Schüler nach einer längeren Phase des Vergessens unvorbereitet Anforderungen an grundlegende Vorstellungen und Kenntnisse sowie Fertigkeiten und Fähigkeiten sicher erfüllen muss. Dies ist ein Ergebnis eines bestimmten Typs von Lernpro-

zessen, die der Herausbildung eines sicheren Wissens und Könnens dienen und mit bestimmten Formen der Unterrichtsgestaltung verbunden sind wie z. B. tägliche Übungen. Analog können auch die übrigen Ebenen mit spezifischen Anforderungssituationen in und nach der Schule und spezifischen Lernprozessen verbunden werden, wie aus der Charakterisierung der Ebenen bereits teilweise ersichtlich ist.

Die Betrachtungen auf der Ebene der sicheren Verfügbarkeit bewegen sich in den Anforderungen, die die Schüler in Klausuren und zentrale Erhebungen zur Bewertung ihrer Leistungen am Ende der Schulzeit zu erbringen haben. Zu dem Zeitpunkt der Leistungserhebung müssen die betreffenden psychischen Dispositionen eine entsprechende Qualität besitzen, um die gestellten Anforderungen erfolgreich zu bewältigen. Im Anschluss an die Überprüfungen sinkt erfahrungsgemäß das erreichte Niveau relativ schnell wieder ab. Es sollte jedoch zu einem späteren Zeitpunkt, etwa im Rahmen der beruflichen Ausbildung möglich sein, durch eine entsprechende Reaktivierung das schon einmal vorhandene Niveau in relativ kurzer Zeit wieder zu erreichen. Dies ist mit Blick auf die nachfolgenden Bildungswege der Schüler als ausreichend anzusehen, da in der Regel die dort benötigten Bestandteile der schulischen Wissens und Könnens direkt oder immanent wiederholt werden.

Auf der Grundlage unserer Erfahrungen bei der Strukturierung von Leistungsdispositionen und mit Leistungserhebungen halten wir es für sinnvoll, möglichst eng begrenzte und kontextbezogene Kompetenzbereiche zu modellieren und schließen uns damit den Auffassungen von Klieme und Leutner (s. S. 124 f.) an. Beispiele für solche Bereiche bzw. so genannte Domänen sind das Wissen und Können im Arbeiten mit Prozentangaben, das Wissen und Können zu den Merkmalen und Eigenschaften geometrischer Körper, das räumliche Vorstellungsvermögen, das Wissen und Können und inhaltlichen und formalen Lösen von Gleichungen und Ungleichungen u. a.

Funktionen des Kompetenzebenenmodells

Das Modell soll eine Orientierungsgrundlage für folgende Forschungen bzw. Überlegungen sein:

1. Forschungen zu Kompetenzmodellen, d. h. zur Struktur und zum Niveau von kognitiven Leistungsdispositionen in Kompetenzbereichen
2. Überlegungen von Lehrern und Fachkollegien zu Schwerpunktsetzungen bei der Unterrichtsplanung und bei der Auswahl von Unterrichtsmethoden

3. Forschungen zur Entwicklung von psychometrischen Modellen und Messverfahren

Diese Funktionen sollen im Folgenden am Beispiel des Wissens und Könnens im Arbeiten mit Größen erläutert werden. Dazu zählen wir folgende Komponenten:

- Kenntnisse zu den Begriffen Größe, Währung, Länge, Masse, Zeit, Flächeninhalt und Volumen und den Einheiten der Größen
- Anschauliche Vorstellungen zu Einheiten der Größen sowie zu bestimmten Teilen und Vielfachen der Einheiten
- Können im Schätzen von Größen
- Fertigkeiten im Umrechnen von Größenangaben
- Kenntnisse zum Rechnen mit Größen

Das Können im Lösen von Sachaufgaben ist ein gesonderter Könnensbereich, der außer speziellen Aufgaben zu Größenvorstellungen bzw. zum Schätzen von Größen nicht zum Arbeiten mit Größen gerechnet wird.

Rolle des Modells bei der Strukturierung und Niveaubestimmung von kognitiven Leistungsdispositionen

Aus der Forderung nach einer hohen Verfügbarkeit und Disponibilität des sicheren Wissens und Könnens ergibt sich die Notwendigkeit einer Beschränkung auf möglichst wenige Anforderungen. Eine solche Auswahl und Beschränkung kann nicht in der Verantwortung eines einzelnen Lehrers liegen sondern nur auf Landes- oder Bundesebene erfolgen. Es ist allerdings zu beachten, dass zum sicheren Wissen und Können nicht nur Kenntnisse und Fertigkeiten sondern auch grundlegende Vorstellung zu zentralen mathematischen Begriffen, Denk- und Arbeitsweisen gehören.

Bei unseren Projekten zum sicheren Wissen und Können zeigte sich, dass es Lehrern schwer fällt, auf dieser Ebene zu denken und dabei die Mehrzahl der üblichen Inhalte des Mathematikunterrichts außerhalb der Betrachtungen zu lassen. Es bestand meist das immanente Bestreben, viel mehr Dinge für unbedingt notwendig zu halten.

Ausgangspunkt der Überlegungen im Projekt zum sicheren Wissen und Können im Arbeiten mit Größen waren die meist wenig befriedigenden Ergebnisse unserer Vergleichsarbeiten in den Klassen 7 und 9 (vgl. CD: Bericht_Vergleichsarbeiten_Mathematik_01_02.pdf, S. 8 ff.). Daraus entstand der Gedanke, bei jeder Größenart eine Beschränkung auf wenige Einheiten vorzunehmen. Dazu haben wir überlegt, welche Bedeutung die einzelnen Einheiten im Alltag haben. Nach längeren Diskussionen haben wir uns darauf verständigt, dass z. B. von den Einheiten für das Volumen die Einheiten

Milliliter, Liter und Kubikmeter die größte Bedeutung besitzen. Auf der Ebene des sicheren Wissens und Könnens sollten deshalb die Schüler Umrechnungen zwischen diesen drei Einheiten jederzeit als sichere Fertigkeit vornehmen können, sie sollten sichere Größenvorstellungen zu diesen Einheiten besitzen und diese bei der Wahl sinnvoller Einheiten und beim Schätzen des Volumens von betreffenden Objekten sicher anwenden können.

Die Überlegungen zu den wesentlichen Merkmalen der Begriffe und ihrer festen Verankerung im semantischen Netz der Schüler führen zu sprachlich-linguistischen Analysen, die nach unseren Erfahrungen eine Reihe von bisher nicht beachteten epistemologischen und sprachlichen Schwierigkeiten hervorbringen. Bei den Größen betraf dies vor allem den Begriff Größe selbst. Im Rahmen der Projekte zum geometrischen Wissen und Können führten diese Analysen bei fast allen zentralen Begriffen zu neuen Einsichten hinsichtlich der betreffenden kognitiven Strukturen.

Bei den Überlegungen zur Ebene der reaktivierbaren Verfügbarkeit zum Arbeiten mit Größen ordneten wir in diese Ebene den Umgang mit allen Einheiten nach dem internationalen Einheitensystem ein. Die Schüler müssen in der Lage sein, nach einer entsprechenden Wiederholung und eventuell unter Zuhilfenahme von Nachschlagewerken alle entsprechenden Größenangaben umrechnen zu können. Sie sollten weiterhin Größenvorstellungen zu allen diesen Einheiten sowie auch zu bestimmten Teilen oder Vielfachen dieser Einheiten besitzen.

Im Rahmen der Betrachtungen zur Ebene der exemplarischen Verfügbarkeit denken wir an solche Bildungsinhalte, die nicht Bestandteil von Abschlussprüfungen sind oder sein können und wo es nur darum geht, erste Spuren und Eindrücke im Gedächtnis zu hinterlassen, die sich auf spätere Lernprozesse auswirken. In Bezug auf das Arbeiten mit Größen geht es um den Umgang mit historischen Einheiten oder zu nichtdezimalen Einheiten wie Barrel, die Kenntnisse zu Vorsilben oder die Vorstellungen zu großen Zahlen wie etwa zu astronomischen Einheiten oder zu großen Geldbeträgen. Dabei sollte auch die Möglichkeit bedacht werden, durch eine besonders beeindruckende Unterrichtsgestaltung Erinnerungen im episodischen Gedächtnis zu erzeugen. So könnte z. B. ein Schülervortrag zu nichtdezimalen Raummaßen gehalten werden oder entsprechende Videos zu diesem Thema gezeigt werden, die die Höhe eines Geldstapels oder die Zeit zum Zählen einer größeren Geldmenge gemessen oder ein maßstäbliches Modell des Sonnensystems hergestellt werden.

Rolle des Modells bei Überlegungen von Lehrern und Fachkollegien zur Schwerpunktsetzung bei der Unterrichtsplanung und bei der Auswahl von Unterrichtsmethoden

Aus den bisherigen Betrachtungen wird bereits deutlich, dass das Modell die Lehrer bei der Wahl ihrer Schwerpunkte und Methoden des Unterrichts orientieren kann. So sollten die als sicheres Wissen und Können ausgewiesenen Dispositionen Hauptinhalt permanenter Wiederholungen z. B. im Rahmen täglicher Übungen sein. Der Grad der Beherrschung sollte durch den selbst bestimmten Einsatz von Testverfahren kontinuierlich überprüft werden.

Das für die Ebene der exemplarischen Verfügbarkeit ausgewählte Wissen und Können sollte den Schülern nur in exemplarischer aber beeindruckender Weise ohne dahinter stehenden Prüfungsdruck nahe gebracht werden.

Mit der Gewichtung von Zielen, die sich aus dem Modell ergibt, wird die Absicht verfolgt, einen neuen Zugang zur Bewältigung des permanenten Stoff-Zeit-Problems anzubieten, indem den Lehren Möglichkeiten aufgezeigt werden, die gegenwärtigen Pläne zu erfüllen, ohne auf größere Inhalts- und Zielbereiche zu verzichten.

Rolle des Modells bei Forschungen zur Entwicklung von psychometrischen Modellen und Messverfahren

Die bisherigen Leistungserhebungen beziehen sich fast alle auf die Gesamtstruktur der psychischen Dispositionen, da das gesamte Spektrum der Anforderungen zugrunde gelegt wird. Erfolgen diese Erhebungen tatsächlich unvorbereitet für die Schüler, so kann man ihnen nicht vorwerfen, dass sie bestimmte Anforderungen nicht aus dem Stand bzw. nur unvollständig erfüllen können. Im Interesse der Schüler werden die Lehrer bei jährlichen Wiederholungen dieser Leistungserhebungen ihre Schüler anhand der bisherigen Anforderungen darauf vorbereiten, womit die Aussagekraft der Ergebnisse wesentlich eingeschränkt wird.

Verfahren zur Kompetenzmessung, die sich an dem Modell der Kompetenzebenen orientieren, müssen die Besonderheiten der zu erfassenden Leistungsdispositionen berücksichtigen, so dass jeweils spezielle Verfahren für die einzelnen Ebenen entwickelt werden müssen. Das Itemuniversum für die einzelnen Kompetenzebenen muss im Ergebnis von Untersuchungen zu dem Wechselverhältnis von gesellschaftlich determinierten Leistungsanforderungen und Leistungsmöglichkeiten der Schule normativ be-

stimmt werden. Damit wäre eine kontentvalide Grundlage (Klauer 1987) für die Messverfahren vorhanden.

Bei Messungen auf der Ebene der reaktivierbaren Verfügbarkeit sollte entsprechend der Qualitätsparameter dieser Ebene den Schülern eine Möglichkeit zur Reaktivierung ihres Wissens und Könnens gegeben werden. Dies könnte zum Beispiel dadurch erfolgen, dass die Schüler vorher ein entsprechendes Material zur Wiederholung lesen können oder auch während des Testes bestimmte Texte mit Informationen und Beispielen verwenden können. Diese Bedingungen würden den üblichen Situationen nach der Schule entsprechen, in denen auch meist Bücher oder Nachschlagewerke verwendet werden, wenn es um anspruchsvolle Aufgaben geht.

Es ist nach unserer Einschätzung auch durchaus möglich, mit speziellen Verfahren das Wissen und Können von Schülern zu überprüfen, das vereinbarungsgemäß nur exemplarisch im Unterricht vermittelt wird. Bei den Verfahren sollte die Möglichkeit vorhanden sein, die in der Klasse realisierten Lernprozesse zu berücksichtigen

Bei der Auswertung der Daten ist die Hierarchie der Ebenen zu beachten, die unteren Ebenen sind jeweils in der darüber liegenden enthalten. Die Tests auf den Ebenen der reaktivierbaren und der exemplarischen Verfügbarkeit sollten den Zuwachs zu den darunter liegenden Ebenen abbilden.

Zur Bestimmung der Lösungswahrscheinlichkeit auf der Ebene der sicheren Verfügbarkeit

Ausgehend von den Vorschlägen von Klauer (1987) plädieren wir dafür, die notwendige Lösungswahrscheinlichkeit auf den einzelnen Ebenen normativ festzusetzen.

Als zu erreichenden Beherrschungsgrad des sicheren Wissens und Könnens einigten wir uns in der Auswertungsgruppe für die Vergleichsarbeiten in Mecklenburg-Vorpommern auf die Marke 67 % (bzw. $\frac{2}{3}$). Ausgangspunkt dieser Festsetzung war die Überlegung, dass es als ausreichend anzusehen wäre, wenn Schüler von drei Aufgaben aus dem Bereich des sicheren Wissens und Könnens immer zwei richtig lösen würde. Über die Interpretationen dieses Wertes von 67 % wurde allerdings keine Übereinstimmung der Ansichten erreicht. Es gibt zwei unterschiedliche Interpretationen, die sich auch in den Auswertungsberichten zu den Vergleichsarbeiten widerspiegeln.

Bei einer der beiden Interpretationen wird die Marke von 67 % als die mindestens zu erreichende durchschnittliche Erfüllungsquote in einer Population bei einer einzelnen Anforderung aufgefasst. Wenn bei einer Aufgabe bzw. einer Teilaufgabe die Erfüllungsquote über 67 % lag, wurde dies als ausreichende Qualität des sicheren Wissens und Könnens angesehen.

Gegen diese Betrachtungsweise kann man zwei Einwände vorbringen. Zum einen werden durch eine solche Durchschnittsbetrachtung mögliche Unterschiede im Leistungsvermögen überdeckt. Wenn im Extremfall zwei Drittel der Schüler die Leistung voll beherrschen und ein Drittel gar nicht, ergibt sich im Schnitt auch ein Wert von etwa 67 %. Ein Grundanliegen des Konzeptes des sicheren Wissens und Könnens besteht aber gerade darin, dass *jeder* Schüler über eine bestimmte Mindestkompetenz verfügt.

Ein weiterer wesentlicher Einwand ergibt sich aus der Betrachtung eines einzelnen Schülers beim Lösen einer geringen Anzahl von mehreren Aufgaben der gleichen Art. Diese Betrachtung entspricht der Situation, wenn nach der Schule etwa in der Berufsausbildung ein Auszubildender mit elementaren Anforderungen konfrontiert wird, die er aus dem Stand zu lösen hat wie z. B. drei einfache Aufgaben zur Prozentrechnung. Die Klagen aus der Berufsausbildung über unzureichende Grundkenntnisse waren ein wesentliches Motiv für die Durchführung unserer Vergleichsarbeiten.

Um dies an Beispielen numerisch zu verdeutlichen, sollen folgende stark vereinfachende Annahme getroffen werden. Der Schüler soll über die betrachtete Fähigkeit zur richtigen Beantwortung eines Items mit einer Wahrscheinlichkeit von $\frac{2}{3}$ verfügen, d. h. im Sinne der ersten Interpretation im Mittel eine Erfüllungsquote von 67 % zu erwarten ist. Weiterhin werde vorausgesetzt, dass alle im Test verwendeten Items aus dem zum Kompetenzbereich gehörenden Itemuniversum stammen und alle für den Schüler den gleichen Schwierigkeitsgrad besitzen. Das Lösen der betreffenden Aufgaben durch den Schüler soll stochastisch voneinander unabhängig sein. Unter diesen Voraussetzungen kann die Wahrscheinlichkeit für die Anzahl richtiger Lösung mit einem Binomialmodell berechnet werden (s. Tabelle 2.2).

Aus der Tabelle ist ersichtlich, dass ein Schüler unter diesen Voraussetzungen von 3 Aufgaben zu einem Thema mit einer Wahrscheinlichkeit von 26 % keine oder nur eine löst bzw. von 6 Aufgaben mit 32%iger Wahrscheinlichkeit nur 3 oder weniger schafft. Dies entspricht nicht den Intensionen eines *sicheren* Wissens und Könnens. In 26 % bzw. 32 % der Fälle würden die Berufsschullehrer der Behauptung widersprechen, dass die Schüler tatsächlich immer 2 von 3 Aufgaben richtig lösen können.

Tabelle 2.2: Wahrscheinlichkeiten für das Lösen von mehreren Aufgaben bei einer Lösungswahrscheinlichkeit von $\frac{2}{3}$ für eine Aufgabe

Anzahl der Items	Der Schüler löst $\frac{1}{3}$ oder weniger Aufgaben richtig		Der Schüler löst weniger als $\frac{2}{3}$ der Aufgaben richtig	
	Anzahlen	Wahrscheinlichkeit ¹¹	Anzahlen	Wahrscheinlichkeit
3	0 bis 1	0,26	0 bis 1	0,26
6	0 bis 2	0,10	0 bis 3	0,32
9	0 bis 3	0,042	0 bis 5	0,35
12	0 bis 4	0,019	0 bis 7	0,37
15	0 bis 5	0,0085	0 bis 9	0,38
18	0 bis 6	0,0039	0 bis 11	0,39
21	0 bis 7	0,0018	0 bis 13	0,40

Bei einer zweiten Interpretation wurde die Marke $\frac{2}{3}$ gedeutet als der Anteil einer Anzahl gleichartiger Items, den jeder Schüler *mindestens* richtig beantworten kann. Hier wird der Wert $\frac{2}{3}$ nicht als Durchschnittswert in einer Population sondern als Eigenschaft eines individuellen Personenmerkmals aufgefasst.

Durch eine fiktive Rechnung unter den gleichen stark vereinfachenden Annahmen wie oben sollen erneut die Probleme und numerischen Dimensionen verdeutlicht werden. Es werde dazu wie vorher vorausgesetzt, dass alle zum Test verwendeten Items aus dem zum Kompetenzbereich gehörenden Itemuniversum stammen, für den Schüler den gleichen Schwierigkeitsgrad besitzen und dass das Lösen der betreffenden Items stochastisch voneinander unabhängig ist.

Als Wahrscheinlichkeit, dass der Schüler weniger als $\frac{2}{3}$ der Aufgaben löst, setzen wir zunächst den oft üblichen Wert von 5 % fest, womit eine Irrtumswahrscheinlichkeit für einen Fehler 1. Art bei einer entsprechenden Hypothese in einem Signifikanztest gegeben ist. Mit einem Binomialmo-

¹¹ Die Wahrscheinlichkeiten werden nur mit zwei wesentlichen Ziffern angegeben.

dell können nun folgende Wahrscheinlichkeiten berechnet werden, wobei X die Anzahl der richtig beantworteten Items ist.

Tabelle 2.3: Wahrscheinlichkeiten für das Lösen von weniger als $\frac{2}{3}$ der Aufgaben für verschiedene Lösungswahrscheinlichkeiten einer Aufgabe

Anzahl der Items	maximale Zahl nicht gelöster Items	minimale Wahrscheinlichkeit ¹² p für $P(X < \frac{2}{3}n) \leq 0,05$	$P(X < \frac{2}{3}n)$	
			$p = 0,85$	$p = 0,80$
3	1	0,87	0,061	0,104
6	3	0,85	0,047	0,099
9	5	0,84	0,034	0,086
12	7	0,82	0,023	0,073
15	9	0,81	0,017	0,061
18	11	0,81	0,012	0,051
21	13	0,80	0,008	0,043

Die Tabelle zeigt, dass die minimale Lösungswahrscheinlichkeit einer einzelnen Aufgabe in Abhängigkeit von der Anzahl der Aufgaben 80 % bis 87 % betragen muss, damit die Hypothese, dass ein Schüler mindestens $\frac{2}{3}$ der gestellten Aufgaben richtig löst, nicht abgelehnt wird.

Geht man umgekehrt von einer Lösungswahrscheinlichkeit für eine einzelne Aufgabe von 80 % aus, so beträgt die Wahrscheinlichkeit, dass von 3 bzw. 6 Aufgaben mindestens 2 bzw. 4 Aufgaben richtig gelöst werden, etwa 90 %.

Damit ergeben sich aus dieser Betrachtungsweise höhere Anforderungen an das auszubildende Niveau der Beherrschung und es wird verständlich, dass von Lehren dies zunächst als unrealistisch abgelehnt wird.

Die Betrachtungen beziehen sich auf einen einzelnen Schüler, es geht also um eine individuelle Diagnostik und Förderung. Dies entspricht eher dem Konzept des sicheren Wissens und Könnens im Unterricht, denn es soll jeder ein minimales Kompetenzniveau besitzen und wer das Niveau bereits erreicht hat, muss nicht weiter daran arbeiten.

¹² Die Wahrscheinlichkeiten werden nur mit zwei wesentlichen Ziffern angegeben.

Ein Nachteil dieser Betrachtungsweise ist, dass die Lösungswahrscheinlichkeit für eine Aufgabe nicht direkt ermittelt, sondern nur aus den Anzahlen richtiger Antworten bei mehreren Items geschätzt werden kann. Die minimal notwendige Lösungswahrscheinlichkeit einer einzelnen Aufgabe ist außerdem von der Anzahl der zu lösenden Aufgaben abhängig. Es muss weiterhin eine Irrtumswahrscheinlichkeit festgelegt werden.

In der Arbeit eines Lehrers bei der statistischen Auswertung von Leistungserhebungen spielen die durchschnittliche Punktzahl bzw. der Zensurdurchschnitt und die Punkt- bzw. Zensurenverteilung die wichtigste Rolle. Bei der ersten Betrachtungsweise wäre für einen Lehrer das Ziel erreicht, wenn im Durchschnitt mindestens 67 % der Punkte erreicht sind. Bei der zweiten Betrachtungsweise müsste er überprüfen, ob jeder Schüler mindestens 67 % der Punkte erreicht hat, in der Arbeit also die Noten 4, 5, und 6 nicht vorkommen. Beide Überprüfungen sind für einen Lehrer in einfacher Weise möglich.

Zusammenfassend plädieren wir für die zweite schülerbezogene Interpretation der Marke 67 % bzw. $\frac{2}{3}$. Das sich daraus implizit eine höhere Lösungswahrscheinlichkeit für eine einzelne Aufgabe ergibt, sollte nicht unbedingt in den Vordergrund gestellt werden. Im Rahmen einer Leistungserhebung sollten für einen bestimmten Bereich des sicheren Wissens und Könnens nur 6, 9 oder 12 Items verwendet werden. Es kann sich dabei um eine tägliche Übung oder einen Teil einer Leistungskontrolle bzw. Klausur handeln. Der Lehrer hat dann nur festzustellen, ob jeder Schüler 4, 6 oder 8 Items richtig beantwortet hat. Für Schüler, für die dies nicht zutrifft, sind dann spezielle Maßnahmen erforderlich.

Es wäre auch ein schrittweises Vorgehen möglich, indem in einem Test mehrere Kompetenzbereiche mit jeweils drei Aufgaben getestet werden. Hat ein Schüler oder ein Auszubildender in einem der Bereiche weniger als zwei Aufgaben richtig gelöst, werden ihm zu diesem Bereich eine größere Anzahl weiterer Aufgaben vorgelegt, so dass dann eine genauere Diagnose seines betreffenden Kompetenzniveaus möglich ist.