Lernerfolgskontrollen im Mathematikunterricht

Detlef Lind

Vorlesung im WS 2005/2006

Bergische Universität Wuppertal

Inhaltsverzeichnis

I Kompetenzen und Lernziele

I.1 Zitate aus Richtlinien

In den Kernlehrplänen für Hauptschulen, Realschulen und die Sekundarstufe I der integrierten Gesamtschulen des Landes NRW werden verbindliche Anforderungen am Ende der Schuljahre 6, 8 und 10 formuliert.
Dort heißt es vorab:

> ***Vorbemerkung: Kernlehrpläne als neue Form der Unterrichtsvorgaben***
>
> *Kernlehrpläne sind ein wichtiges Element eines zeitgemäßen und umfassenden Gesamtkonzepts für die Entwicklung und Sicherung der Qualität schulischer Arbeit. Sie sind im Zusammenhang zu sehen mit den Lernstandserhebungen, die in Nordrhein- Westfalen 2004 zum ersten Mal in den Klassen 9 der Sekundarstufe I durchgeführt werden, und mit den landeseinheitlichen Abschlussprüfungen am Ende der Klasse 10 ab 2007.*
>
> *Kernlehrpläne*
>
> - *sind standardorientierte Lehrpläne, in denen die erwarteten Lernergebnisse als verbindliche Bildungsstandards im Mittelpunkt stehen.*
> - *beschreiben die erwarteten Lernergebnisse in der Form von fachbezogenen Kompetenzen, die fachdidaktisch begründeten Kompetenzbereichen zugeordnet sind.*
> - *zeigen, in welchen Stufungen diese Kompetenzen im Unterricht der Klassen 5 bis 10 erreicht werden können, indem sie die erwarteten Kompetenzen am Ende der Klassen 6, 8 und 10 bezeichnen.*
> - *beschränken sich dabei auf wesentliche Kenntnisse und Fähigkeiten und die mit ihnen verbundenen Inhalte und Themen, die für den weiteren Bildungsweg unverzichtbar sind und die den Lehrerinnen und Lehrern aus ihrer bisherigen Unterrichtspraxis im Wesentlichen bekannt sind.*
> - *bestimmen durch die Ausweisung von verbindlichen Erwartungen die Bezugspunkte für die Überprüfung der Lernergebnisse und der erreichten Leistungsstände in der schulischen Leistungsbewertung, den Lernstandserhebungen und den Abschlussprüfungen mit zentral gestellten Aufgaben für die schriftlichen Prüfungen.*
> - *schaffen so die Voraussetzungen, um definierte Anspruchsniveaus an der Einzelschule und im Land zu sichern.*

Indem Kernlehrpläne sich auf die zentralen Kompetenzen beschränken, geben sie den Schulen die Möglichkeit, sich auf diese zu konzentrieren und ihre Beherrschung zu sichern. Die Schulen können dabei entstehenden Freiräume zur Vertiefung und Erweiterung der behandelten Unterrichtsinhalte und damit zu einer inhaltlichen und thematischen Profilbildung nutzen.
Die Bildungsstandards der Kultusministerkonferenz sind auf den mittleren Schulabschluss bezogen und insofern schulformübergreifend angelegt, um für den gleichen Abschluss ein einheitliches Mindestniveau zu sichern. Die Kernlehrpläne greifen die in den KMK-Standards enthaltenen schulformübergreifenden Ansprüche auf und berücksichtigen gleichzeitig die Besonderheiten der einzelnen Schulformen und Bildungsgänge. Diesen wird in der Beschreibung der Standards und in der Art des methodischen Zugriffs Rechnung getragen. Beispielhafte Aufgabenstellungen im Bildungsserver learn-line verdeutlichen die konkreten, zum Teil unterschiedlichen Kompetenzerwartungen (`www.learn-line.nrw.de/angebote/kernlehrplaene`).
Die bisherigen Richtlinien der Schulformen bleiben bis auf weiteres in Kraft. Sie beschreiben die Aufgaben und Ziele der Schulformen in der Sekundarstufe I und enthalten auch die spezifischen Hinweise zum Lehren und Lernen in diesen Schulformen. Die vorgelegten Kernlehrpläne und die in ihnen enthaltenen Standards stellen einen Einstieg in eine längerfristige Entwicklung dar. Die in den Kernlehrplänen enthaltenen Kompetenzbeschreibungen beziehen sich wie die in den Bildungsstandards der KMK vorerst auf ein mittleres Anspruchsniveau (Regelstandards). Perspektivisch sollen sowohl für die KMK-Bildungsstandards wie für die Bildungsstandards in den Kernlehrplänen Kompetenzstufen auf der Basis empirisch und fachdidaktisch geklärter Kompetenzstufenmodelle ausgewiesen werden. Auf dieser Basis können dann das angestrebte Mindestniveau (Mindeststandards), der Regelfall und ein Exzellenzniveau ausgewiesen werden. Die Kultusministerkonferenz hat dazu ein wissenschaftliches Institut gegründet, das solche Kompetenzstufen im Laufe der nächsten Jahre entwickeln wird. Die landeseigenen Lernstandserhebungen werden hierzu ebenfalls Hinweise geben.

Es sollen nun Auszüge aus dem Kernlehrplan der IGS folgen, da in diesem die verschiedenen Leistungsniveaus angesprochen werden. Für stärker Interessierte hat der Dozent die Kernlehrpläne für Realschulen und die Sekundarstufe I der integrierten Gesamtschulen zusätzlich zum Skript ins Netz gestellt.

1 Aufgaben und Ziele des Mathematikunterrichts

Schülerinnen und Schüler sollen im Mathematikunterricht der Sekundarstufe I

- *Erscheinungen aus Natur, Gesellschaft und Kultur mit Hilfe der Mathematik wahrnehmen und verstehen (Mathematik als Anwendung)*
- *mathematische Gegenstände und Sachverhalte, repräsentiert in Sprache, Symbolen und Bildern, als geistige Schöpfungen verstehen und weiterentwickeln (Mathematik als Struktur)*
- *in der Auseinandersetzung mit mathematischen Fragestellungen auch überfachliche Kompetenzen erwerben und einsetzen (Mathematik als kreatives und*

intellektuelles Handlungsfeld). Hierbei erkennen sie, dass Mathematik eine historisch gewachsene Kulturleistung darstellt. Zugleich erleben sie Mathematik als intellektuelle Herausforderung und als Möglichkeit zur individuellen Selbstentfaltung und gesellschaftlichen Teilhabe. Sie entwickeln personale und soziale Kompetenzen, indem sie lernen,

- *gemeinsam mit anderen mathematisches Wissen zu entwickeln und Probleme zu lösen (Kooperationsfähigkeit als Voraussetzung für gesellschaftliche Mitgestaltung).*

- *Verantwortung für das eigene Lernen zu übernehmen und bewusst Lernstrategien einzusetzen (selbstgesteuertes Lernen als Voraussetzung für lebenslanges Lernen).*

Mathematische Grundbildung umfasst die Fähigkeit, die Rolle zu erkennen, die Mathematik in der Welt spielt, mathematisches Wissen funktional, flexibel und mit Einsicht zur Bearbeitung vielfältiger kontextbezogener Probleme einzusetzen und begründete mathematische Urteile abzugeben. Sie beinhaltet insbesondere die Kompetenz des problemlösenden Arbeitens in inner- und außermathematischen Kontexten. Grundlegend dafür ist die Fähigkeit, komplexe Probleme zu strukturieren sowie reale Probleme in geeigneter Weise mathematisch zu beschreiben, also Modelle zu bilden und zu nutzen. Ebenso gehört zur mathematischen Grundbildung die Fähigkeit, mit anderen über mathematische Fragestellungen zu kommunizieren, d.h. eigene Ideen zu präsentieren und zu begründen sowie die Argumente anderer aufzunehmen.

Diese Kompetenzen bilden sich bei der aktiven Auseinandersetzung mit konkreten Fragestellungen aus den Kernbereichen des Faches Mathematik heraus: Die Mathematik erfasst ebene und räumliche Gebilde mit Mitteln der Geometrie. Für die Operationen mit Zahlen in der Arithmetik hat die Mathematik die Formelsprache der Algebra entwickelt, mit der sich Gesetzmäßigkeiten des Zahlenrechnens darstellen und flexibel nutzen lassen. Zu den Leistungen der Mathematik gehört ferner, dass sie sowohl systematische Abhängigkeiten von Zahlen und Größen mit dem Begriff der Funktion, aber auch zufällige Ereignisse mit dem Begriff der Wahrscheinlichkeit beschreiben kann.

***Mathematische Grundbildung** zeigt sich also im Zusammenspiel von Kompetenzen, die sich auf mathematische Prozesse beziehen und solchen, die auf mathematische Inhalte ausgerichtet sind. Prozessbezogene Kompetenzen, wie z.B. das Problemlö- sen oder das Modellieren werden immer nur bei der Beschäftigung mit konkreten Lerninhalten, also unter Nutzung inhaltsbezogener Kompetenzen erworben und weiterentwickelt.*

fachbezogene Kompetenzen			
prozessbezogene Kompetenzen		**inhaltsbezogene Kompetenzen**	
Argumentieren/ Kommunizieren	kommunizieren, präsentieren und argumentieren	Arithmetik/ Algebra	mit Zahlen und Symbolen umgehen
Problemlösen	Probleme erfassen, erkunden und lösen	Funktionen	Beziehungen und Veränderung beschreiben und erkunden
Modellieren	Modelle erstellen und nutzen	Geometrie	ebene und räumliche Strukturen nach Maß und Form erfassen
Werkzeuge	Medien und Werkzeuge verwenden	Stochastik	mit Daten und Zufall arbeiten

Fig. 1

Die hier genannten Bereiche mathematischer Kompetenzen werden im Folgenden konkretisiert durch eine Beschreibung von Anforderungen am Ende der Sekundarstufe I sowie durch eine Darstellung von Kompetenzerwartungen am Ende der jeweiligen Jahrgangsstufen. Diese Kernkompetenzen sollen Schülerinnen und Schüler nachhaltig und nachweislich erworben haben.

Die inhaltliche und methodische Gestaltung eines Unterrichts, in dem Schülerinnen und Schüler eine solche mathematische Grundbildung erwerben können, ist als Gesamtaufgabe aufzufassen. Inhalte und Methoden des Unterrichts sind eng aufeinander bezogen. Eine einseitig kleinschrittige Methodik, die entlang einer vorgegebenen Stoffsystematik eine Engführung der Lernenden betreibt, ist nicht geeignet, junge Menschen verständnisorientiert in mathematisches Denken einzuführen. Der Unterricht soll Schülerinnen und Schüler bei der Auseinandersetzung mit Mathematik unterstützen. Er soll hierzu eine breite Palette unterschiedlichster Unterrichtsformen aufweisen, die von einer lehrerbezogenen Wissensvermittlung bis hin zu einer selbstständigen Erarbeitung neuer Inhalte reicht. Zudem darf er sich nicht auf die nachvollziehende Anwendung von Verfahren und Kalkülen beschränken, sondern muss in komplexen Problemkontexten entdeckendes und nacherfindendes Lernen ermöglichen. Er sollte inner- und außermathematische Fragestellungen vernetzen und sich dabei an zentralen mathematischen Ideen (Zahl, Messen, räumliches Strukturieren, Algorithmus, Zufall) orientieren. Dieses Vorgehen erlaubt es auch, sich im Unterricht auf Wesentliches zu konzentrieren, ausgewählte Inhalte zu vertiefen und nach dem Prinzip der integrierenden Wiederholung bereits erworbene Kenntnisse und Fähigkeiten zu festigen und zu vertiefen.

2 Anforderungen am Ende der Sekundarstufe I

Für das Ende der Sekundarstufe I werden im Folgenden die Kompetenzen ausgewiesen, die alle Schülerinnen und Schüler erworben haben, die mit Erfolg am Mathematikunterricht teilgenommen haben. Die Schülerinnen und Schüler sollen in der Lage sein, diese Kompetenzen für ihre persönliche Lebensgestaltung, für ihren weiteren Bildungsweg und für ihr berufliches Leben zu nutzen.

*Diese für den Mathematikunterricht in Nordrhein-Westfalen verbindlichen Kompetenzen werden in enger Anlehnung an die Bildungsstandards der KMK auf der Anforderungsebene des **mittleren Schulabschlusses (Fachoberschulreife)** beschrie-*

ben. Hierdurch soll die Vergleichbarkeit der fachlichen Anforderungen für diesen Abschluss in allen Schulformen der Sekundarstufe I gesichert werden.
Zum Erwerb des Qualifikationsvermerks für den Eintritt in die gymnasiale Oberstufe ist Folgendes festzustellen: Der Mathematikunterricht an Gesamtschulen ermöglicht Schülerinnen und Schülern im oberen Leistungsbereich die Fortsetzung des Bildungsganges in der Sekundarstufe II auch bis zum Abitur. Die für den mittleren Schulabschluss (Fachoberschulreife) geforderten Kompetenzen sind in unterschiedlichem Umfang und auf unterschiedlichem Niveau erreichbar. Von Schülerinnen und Schülern, die den Qualifikationsvermerk für den Eintritt in die gymnasiale Oberstufe erwerben, wird erwartet, dass sie die Kompetenzen auf einem höheren Niveau erreichen. Es gibt allerdings für den Qualifikationsvermerk keine curriculare, inhaltliche Definition. Der Vermerk wird vielmehr auf Grund des Notenbildes in der Versetzungskonferenz vergeben. Entsprechende fachliche Kompetenzen werden daher auch nicht gesondert ausgewiesen.
An der Gesamtschule erwerben Schülerinnen und Schüler auch den ***Hauptschulabschluss nach Klasse 10****. Im Vergleich zu dem unten aufgeführten Kompetenzprofil für den mittleren Schulabschluss (Fachoberschulreife) sind die Anforderungen an diese Schülerinnen und Schüler in Umfang und Anforderungshöhe insgesamt geringer. Diejenigen Schülerinnen und Schüler, die den* ***Hauptschulabschluss*** *mit Vollendung der Vollzeitschulpflicht erwerben, können dabei schon aufgrund der kürzeren Unterrichtszeit nur Teile der in … detaillierter beschriebenen Kompetenzen erreichen. Die Schülerinnen und Schüler, die an der Gesamtschule im Jahrgang 10 den Grundkurs besucht haben, verfügen über die im Folgenden nicht kursiv gesetzten Kompetenzen. Die* ***kursiv-fett*** *gesetzten Textpassagen beschreiben die Kompetenzen, die im Erweiterungskurs zusätzlich erreicht werden müssen.*

Argumentieren/Kommunizieren

kommunizieren, präsentieren und argumentieren

Schülerinnen und Schüler teilen mathematische Sachverhalte zutreffend und verständlich mit und nutzen sie als Begründung für Behauptungen und Schlussfolgerungen.

- Sie entnehmen mathematische Informationen aus Texten, Bildern und Tabellen (Lesekompetenz), ***analysieren und beurteilen die Aussagen***.
- Sie erläutern mathematische Einsichten und Lösungswege mit eigenen Worten und geeigneten Fachbegriffen und präsentieren Überlegungen in kurzen, vorbereiteten Beiträgen ***sowie Problembearbeitungen in vorbereiteten Vorträgen***.
- Sie vernetzen Begriffe, indem sie Beziehungen zwischen Begriffen auch aus verschiedenen Bereichen herstellen, Beispiele angeben und Ober- und Unterbegriffe zuordnen.
- Sie nutzen verschiedene Arten des Begründens und Überprüfens (Plausibilität, Beispiele, ***Argumentationsketten***).

- Sie vergleichen Lösungswege und Darstellungen, ***überprüfen und bewerten Problembearbeitungen***.

Problemlösen

Probleme erfassen, erkunden und lösen

Schülerinnen und Schüler strukturieren und lösen inner- oder außermathematische Problemsituationen, in denen ein Lösungsweg nicht unmittelbar erkennbar ist bzw. bei denen nicht unmittelbar auf erlernte Verfahren zurückgegriffen werden kann.

- Sie geben inner- und außermathematische Problemstellungen mit eigenen Worten wieder, erkunden sie, stellen Vermutungen auf und zerlegen Probleme in Teilprobleme.
- Sie nutzen verschiedene Darstellungsformen, mathematische Verfahren und nutzen Problemlösestrategien wie Überschlagen, Beispiele finden, systematisches Probieren, Schlussfolgern, Zurückführen auf Bekanntes und Verallgemeinern.
- Sie überprüfen ***und bewerten*** Lösungswege und Ergebnisse, ***auch die Möglichkeit mehrerer Lösungen***.

Modellieren

Modelle erstellen und nutzen

Schülerinnen und Schüler nutzen Mathematik als Werkzeug zum Erfassen von Phänomenen der realen Welt.

- Sie übersetzen Realsituationen in mathematische Modelle (Terme, Gleichungen, Funktionen, Figuren, Diagramme, Tabellen, Zufallsversuche) und ordnen mathematischen Modellen passende Realsituationen zu.
- Sie überprüfen und interpretieren die im mathematischen Modell gewonnene Lösung in der jeweiligen realen Situation, ***bewerten*** und verändern gegebenenfalls ihren Lösungsweg oder das Modell.

Werkzeuge

Medien und Werkzeuge verwenden

Schülerinnen und Schüler setzen klassische mathematische Werkzeuge und elektronische Werkzeuge und Medien situationsangemessen ein (Medienkompetenz).

- Sie verwenden Lineal, Geodreieck und Zirkel zum Messen, genauen Zeichnen und Konstruieren.
- Sie nutzen Bücher und das Internet zur Informationsbeschaffung, dokumentieren eigene Arbeitsschritte in schriftlicher Form und verwenden unter anderem Tafel, Folien und Plakate zur Ergebnispräsentation.
- Sie setzen situationsangemessen den Taschenrechner ein und nutzen Geometriesoftware, Tabellenkalkulation und Funktionenplotter zum Erkunden inner- und außermathematischer Zusammenhänge.

Arithmetik/Algebra

mit Zahlen und Symbolen umgehen

Schülerinnen und Schüler besitzen einen Begriff von Zahlen, Größen und ihren Darstellungen, operieren sicher mit ihnen und verwenden die Symbolsprache der Mathematik sachgerecht.

- Sie verwenden Zahlen je nach Situation in unterschiedlichen Darstellungsformen (als Bruch, Dezimalzahl, Prozentzahl und in Zehnerpotenzschreibweise), ordnen und vergleichen sie.
- Sie rechnen mit rationalen und ***irrationalen Zahlen***, nutzen Rechengesetze und systematisches Zählen.
- Sie arbeiten in Anwendungszusammenhängen sachgerecht mit Zahlen, Größen und Variablen und führen Schätzungen und Näherungsrechnungen durch.
- Sie lösen lineare Gleichungen ***und Gleichungssysteme***, quadratische ***und einfache exponentielle*** Gleichungen rechnerisch, grafisch oder durch Probieren.

Funktionen

Beziehungen und Veränderung beschreiben und erkunden

Schülerinnen und Schüler besitzen ein grundlegendes Verständnis von funktionaler Abhängigkeit und nutzen ihre Kenntnisse zum Erfassen und Beschreiben von Beziehungen und Veränderungen in Mathematik und Umwelt.

- Sie stellen funktionale Zusammenhänge, insbesondere lineare, quadratische, ***exponentielle*** Funktionen, Sinusfunktion, in sprachlicher Form, in Tabellen, als Grafen und in Termen dar und interpretieren sie situationsgerecht.
- Sie identifizieren proportionale und antiproportionale Funktionen, wenden Dreisatz, Prozentrechnung und Zinsrechnung an und rechnen mit Maßstäben.
- Sie grenzen lineares, quadratisches und exponentielles Wachstum an Beispielen voneinander ab.

Geometrie

ebene und räumliche Strukturen nach Maß und Form erfassen

Schülerinnen und Schüler erfassen Formen der Ebene und des Raumes und ihre Beziehungen in mathematischen Zusammenhängen sowie in der beobachteten Wirklichkeit und charakterisieren sie anhand ihrer grundlegenden Eigenschaften.

- Sie beschreiben ebene Figuren (Vielecke, Kreise) und Körper (Prismen, Zylinder, Kugeln, Kegel, Pyramiden), Lagebeziehungen und grundlegende Symmetrien mit angemessenen Fachbegriffen und identifizieren sie in ihrer Umwelt.
- Sie zeichnen und konstruieren ebene geometrische Figuren (auch im Koordinatensystem), skizzieren Schrägbilder, entwerfen Netze von Körpern und stellen Körpermodelle her.

- Sie schätzen und bestimmen Winkel, Längen, Flächeninhalte, Oberflächen und Volumina.
- Sie berechnen Größen und begründen Eigenschaften von Figuren mit Hilfe von Symmetrie, einfachen Winkelsätzen, Kongruenz, ***Ähnlichkeit, trigonometrischen Beziehungen, dem Satz des Thales*** und dem Satz des Pythagoras.

Stochastik
mit Daten und Zufall arbeiten

Schülerinnen und Schüler erheben statistische Daten und werten sie aus. Sie beschreiben und beurteilen zufällige Ereignisse mit mathematischen Mitteln.

- Sie planen statistische Erhebungen, nutzen Methoden der Erfassung und Darstellung von Daten (Säulen- und Kreisdiagramme, ***Boxplots***) und bewerten Darstellungen kritisch.
- Sie bestimmen relative Häufigkeiten, Mittelwerte (arithmetisches Mittel, Median) ***und Streumaße (Spannweite, Quartil)*** und interpretieren diese.
- Sie bestimmen Wahrscheinlichkeiten mit Hilfe von Laplace-Regel, ***Baumdiagrammen und Pfadregeln***, nutzen Häufigkeiten zum Schätzen von Wahrscheinlichkeiten und Wahrscheinlichkeiten zur Vorhersage von Häufigkeiten.

Die schuleigenen Lehrpläne und die Evaluation von Unterricht und Unterrichtsergebnissen sind an den oben stehenden Kompetenzprofilen auszurichten.

3 Kompetenzerwartungen am Ende der Jahrgangsstufen 6, 8 und 10

Im Folgenden werden Kompetenzen benannt, die Schülerinnen und Schüler am Ende der Jahrgangsstufen 6, 8 und 10 nachhaltig und nachweislich erworben haben sollen. Sie legen die Art der fachlichen Anforderungen fest. Die Anforderungshöhe und der Komplexitätsgrad der fachlichen Anforderungen sind sowohl im Unterricht als auch in der Leistungsbewertung altersgemäß und mit Bezug auf die Anforderungen der Schulform zu konkretisieren. Kapitel … erläutert die Anforderungen an ausgewählten Muster- und Modellaufgaben.
Die hier benannten Kompetenzen gliedern sich nach den Bereichen des Faches und beschreiben dessen Kern. Sie bauen auf den in der Grundschule erworbenen Kompetenzen auf und machen eine Progression über die Jahrgangsstufen hinweg deutlich. Der Unterricht ist nicht allein auf den Erwerb dieser Kernkompetenzen beschränkt, sondern soll es Schülerinnen und Schülern ermöglichen, auf vielfältige Weise darüber hinausgehende Kompetenzen zu erwerben, weiterzuentwickeln und zu nutzen.
Kompetenzen werden im Unterricht nicht einzeln und isoliert erworben, sondern in wechselnden und miteinander verknüpften Kontexten. Der Unterricht muss dazu vielfältige, die Jahrgangsstufen durchziehende Lerngelegenheiten anbieten. Eine thematisch-inhaltliche Reihenfolge innerhalb der Jahrgangsstufen ist durch den Kernlehrplan nicht festgeschrieben.
Der Kernlehrplan bildet damit einerseits die verpflichtende Grundlage für die Überarbeitung der schuleigenen Lehrpläne. Andererseits eröffnet er Lehrerinnen

und Lehrern weitgehende Freiheiten für die inhaltliche, thematische und methodische Gestaltung von Unterrichtsabläufen. Sie können Schwerpunkte setzen, thematische Vertiefungen und Erweiterungen vornehmen und dabei die Bedingungen der eigenen Schule und der jeweiligen Lerngruppe berücksichtigen.
Im Folgenden werden die fachbezogenen Kompetenzen getrennt nach prozessbezogenen und inhaltsbezogenen Kompetenzen ausgewiesen. Die prozessbezogenen Kompetenzen werden von Schülerinnen und Schülern jedoch immer nur in der Auseinandersetzung mit mathematischen Inhalten erworben. Umgekehrt können sich inhaltsbezogene Kompetenzen nur entfalten, wenn Schülerinnen und Schüler prozessbezogene Kompetenzen aktivieren können. Mathematische Grundbildung zeigt sich in der flexiblen und vernetzten Nutzung dieser prozessbezogenen und inhaltsbezogenen Kompetenzen. Beide Bereiche müssen somit Gegenstand des Unterrichts und der Leistungsbewertung sein.
Zusätzliche Kompetenzen für den Erweiterungskurs werden im Folgenden kursiv-fett gedruckt.

… (es folgen viele Tabellen zu Kompentenzerwartungen, danach die folgende Gesamtübersicht:)

	Arithmetik/Algebra	Funktionen	Geometrie	Stochastik
5/6	• Rechnen mit natürlichen Zahlen, endlichen Dezimalzahlen und einfachen Brüchen • Größen • Ordnen, Vergleichen, Runden • Zahlengerade • Rechenvorteile, systematisches Zählen, Teiler und Vielfache	• Tabellen und Diagramme • Maßstab	• ebene Figuren • Umfang und Fläche von Rechtecken • Quader und Würfel	• Ur- und Strichlisten • Häufigkeitstabellen, Säulendiagramme • arithmetisches Mittel
7/8	• Rechnen mit rationalen Zahlen • lineare Gleichungen	• Wertetabellen, Grafen und Terme • proportionale und antiproportionale Zuordnungen • ***lineare Funktionen*** • Prozentrechnung	• Zeichnen von Dreiecken • Umfang und Fläche von Dreiecken und Vierecken • Schrägbilder, Netze, Körpermodelle • Oberfläche und Volumen • ***einfache Winkelsätze*** • ***Kongruenz***	• Planung und Durchführung von Erhebungen • ***Kreisdiagramme*** • ***Häufigkeit und Wahrscheinlichkeit*** • einstufige Zufallsexperimente
9/10	• Potenzieren, Radizieren • Zehnerpotenzschreibweise • ***Termumformungen*** • ***lineare Gleichungssysteme*** • ***quadratische Gleichungen***	• ***quadratische Funktionen*** • ***Sinusfunktion*** • ***lineares, quadratisches und exponentielles Wachstum (Zinseszins)*** • Zinsrechnung	• Kreisberechnung • ***Dreiecksberechnungen*** • Zylinder, Pyramiden, Kegel, Kugeln • ***Prismen*** • Vergrößern, Verkleinern, Ähnlichkeit • Satz des Pythagoras	• Analyse von grafischen Darstellungen • ***zweistufige Zufallsexperimente*** • ***Pfadregeln*** • ***Boxplots***

Die kursiv und fett gedruckten Kompetenzen beziehen sich am Ende der Jahrgangsstufe 8 auf den Erweiterungskurs, am Ende der Jahrgangsstufe 10 auf den B-Typ.

Fig. 2

Offensichtlich wird in den zitierten Texten die Bezeichnung *Kompetenz* im Sinne von *Können* bzw. *Fähigkeit* verwendet. Bei allen Passagen des Typs

Die Schülerinnen und Schüler … (Handlungsbeschreibung) …

könnte man formulieren:

Es wird erwartet, dass die Schülerinnen und Schüler …

Damit hat man es jedoch eigentlich wieder mit einem Begriff zu tun, der nach einigen Übertreibungen in den 70ern in Misskredit geraten war und beinahe als verpönt galt, – dem **Lernzielbegriff !!!**

I.2 Operationale Lernziele im Mathematikunterricht

Behavioristische Ansätze. Die Lernzieldiskussion vor 30 Jahren unterschied zwischen *allgemeinen* Lernzielen, wie „der Schüler soll lernen, sich kreativ zu verhalten“ und sogenannten *Feinlernzielen*, wie „der Schüler soll den Quotienten zweier Brüche berechnen können“.

Solange es um die Vermittlung von Techniken geht, ist der zweite Begriff nach wie vor brauchbar. Seine Präzisierung in den 60er Jahren war noch stark behavioristisch geprägt[1]:

Sieht man *Lernen* als nur als Verhaltensänderung an, so lässt sich ein Lernziel durch eine Menge $\mathcal{S}$ von Situationen und eine Menge $\mathcal{R}$ von beobachtbaren Reaktionen beschreiben. Dabei sind jeder Situation S (dem sogenannten „Stimulus“) eine (oder mehrere) erwünschte Reaktion(en) R (die sogenannte „Response“) zugordnet.

Schematisch sieht das so aus:

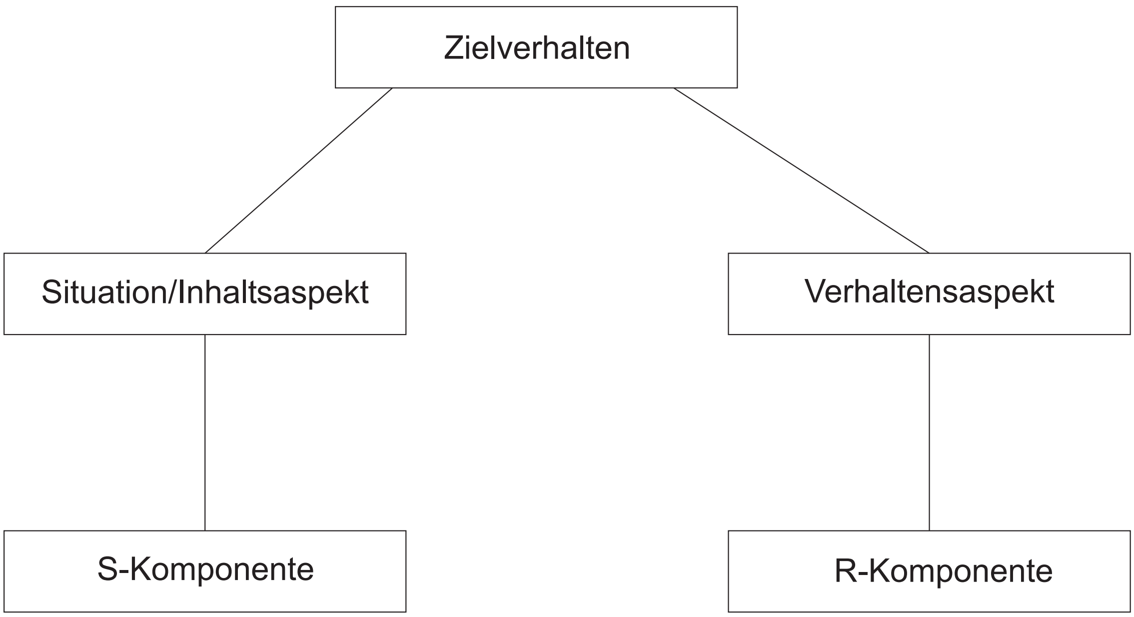


Fig. 3

Beispiel 1:

Das Lernziel „die Schüler(innen) sollen quadratische Gleichungen lösen können“ ließe sich in der Form

[1] vgl. MAGER, R. F.(1965), Lernziele und programmierter Unterricht

$\mathcal{S}$	$\mathcal{R}$
Dem/der Schüler(in) werden 10 Aufgaben der Art Bestimme alle Lösungen der Gleichung $ax^2 + bx + c = 0$ mit ganzzahligem a, b, c und $a \neq o$ nacheinander vorgelegt. Er wird zur Bearbeitung aufgefordert. Erlaubt sind Schreibzeug, leere Blätter und ein Taschenrechner.	Der/die Schüler(in) soll bei mindestens 8 dieser Aufgaben alle Lösungen angeben können.

Damit ist zwar recht genau festgelegt, wann man als Prüfer den Lernerfolg attestiert, es bleibt aber der Zweifel an der willkürlich scheinenden Mindestgrenze 8.

Dass man nicht 10 gelöste Aufgaben verlangt, scheint wohl noch am ehesten gerechtfertigt, da Menschen keine „Roboter“ sind, die auf jeden Knopfdruck ein „erlerntes“ Verhalten zeigen.

Überlegungen dieser Art bewogen K. J. KLAUER 1972 dazu, die Rechtfertigung von Mindestaufgabenzahlen auf den Begriff der individuellen *Lösungswahrscheinlichkeit* zu gründen[2].

Lösungswahrscheinlichkeiten. K. J. KLAUER propagierte den Ansatz, dichotom formulierbare Lernziele (z. B. bei Aufgaben eines bestimmten Typs lediglich die „richtige Bearbeitung“ zu verlangen und daher ihre Bearbeitung durch Schüler(innen) nur nach den Kategorien *richtig* bzw. *falsch* zu bewerten) in folgender Form zu operationalisieren:

Nach Festlegung einer (kleinen!) Risikowahrscheinlichkeit α für das mögliche Scheitern an einer Aufgabe modelliert man die Bearbeitung einer Aufgabe i durch einen Schüler durch ein Zufallsexperiment mit den möglichen Ausgängen 0 (=Misserfolg) und 1 (=Erfolg) und nennt die Erfolgswahrscheinlichkeit des Schülers die *Lösungswahrscheinlichkeit p_i des Schülers* für die Aufgabe i:

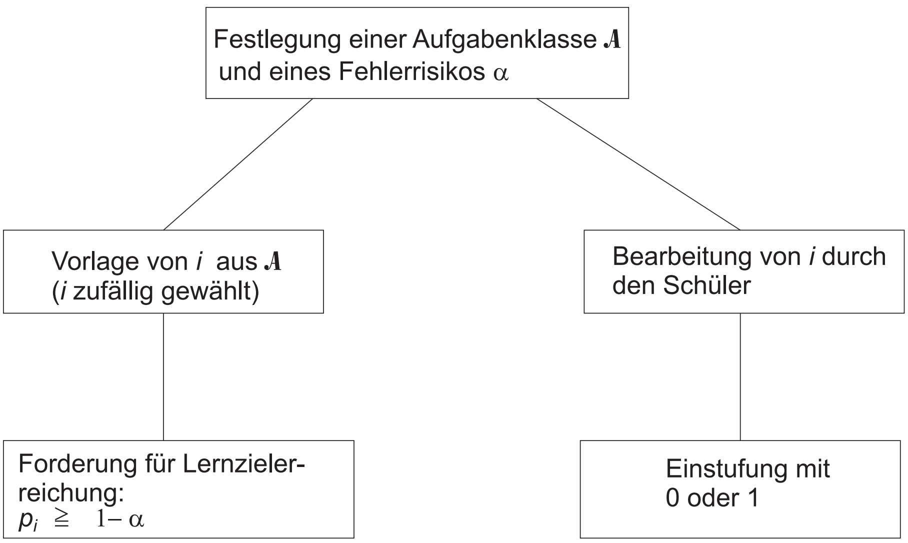


Fig. 4

[2] KLAUER et al. (1972), Lernzielorientierte Tests

Da man bei jeder Aufgabe nur die tatsächlichen Ausgänge beobachten kann, ist man bei einer Lernzielprüfung eigentlich auf die mehrmalige Vorlage von Aufgaben angewiesen.

Wenn α klein gewählt war, scheint eine zufällig gewählte Aufgabe zu reichen, da ein *fiktiver* Lernzielerreicher *a* sie mit mindestens der Wahrscheinlichkeit $1-\alpha$ löst. Das Fehleinstufungsrisiko für *a* beträgt daher *höchstens* α. Dafür wird jedoch ein nicht allzu weit unter der theoretischen Grenze liegender Prüfling *b* mit recht hoher Wahrscheinlichkeit irrtümlich als Lernzielerreicher eingestuft:

Fehlbeurteilungsrisken beim Testen mittels einer einzigen Aufgabe ($\alpha = \mathbf{0,05}$)

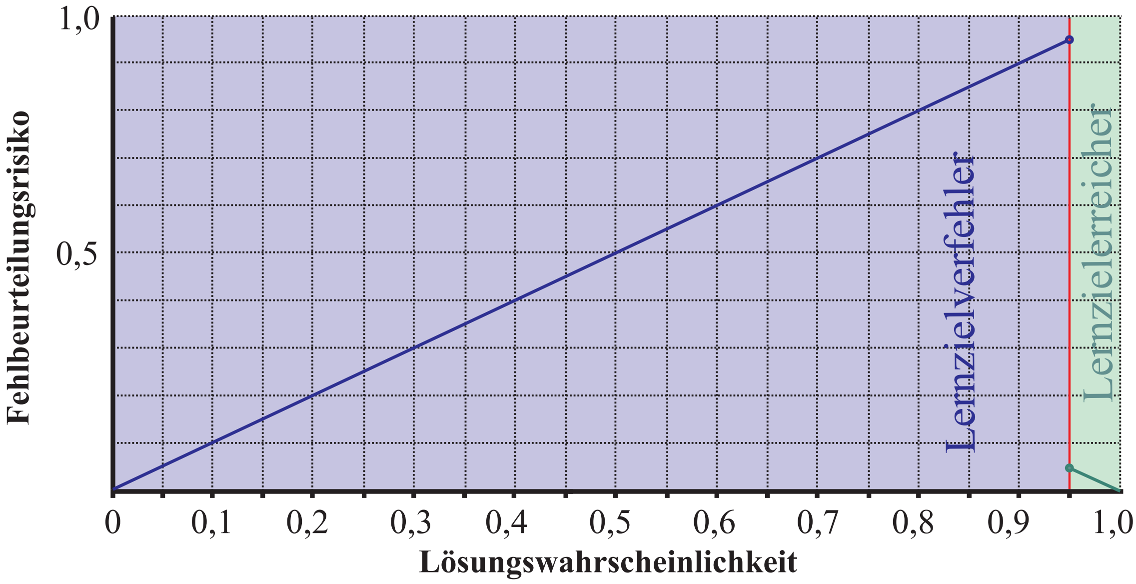


Fig. 5

Wir zeigen zunächst in einer Grafik, wie die Risiken sich ändern, wenn wie in Beispiel 1 für die „erreicht'Einstufung mindestens 8 von 10 Erfolgen verlangt werden:

Fehlbeurteilungsrisken beim Testen (8 von 10 Aufgaben verlangt, $\alpha = \mathbf{0,05}$)

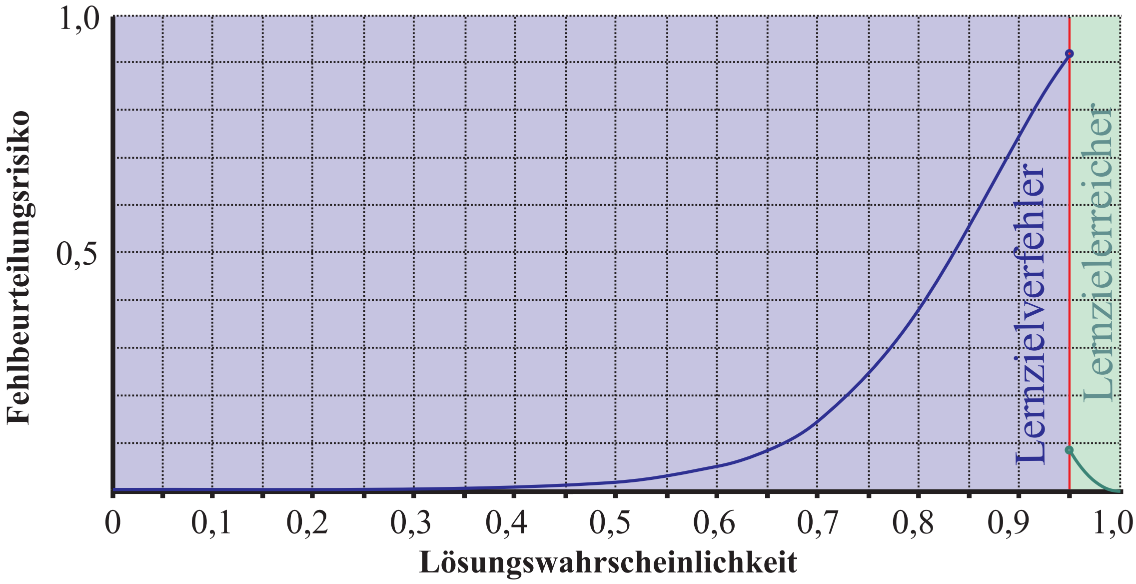


Fig. 6

Solche Darstellungen lassen sich heute komfortabel mit Tabellenkalkulationen berechnen (vgl. die Datei `binomial.xls` im Material zur Vorlesung).
Zum wirklichen Verständnis sind allerdings Kenntnisse in der elementaren Wahrscheinlichkeitsrechnung nötig, die sich über das Durcharbeiten der Datei `WR.pdf` erwerben lassen.
Offensichtlich kann man nicht das Fehlbeurteilungsrisiko für fiktive Lernzielverfehler senken ohne gleichzeitig das entsprechende Risiko für fiktive Lernzielerreicher zu *erhöhen* – und umgekehrt. Es scheint daher angemessener, bei sanktionsfreien Prüfungen diejenige Lösungswahrscheinlichkeit p_0 als Schwelle zwischen „Verfehlern“ und „Erreichern“ festzusetzen, bei derdie Erfolgswahrscheinlichkeit bei einem Prüfungsverfahren des Typs „mindestens k von n“ Aufgaben sind zu lösen genau 50 % beträgt.
So wäre z.B. bei der Forderung „3 oder 4 von insgesamt 4 Aufgaben sind zu lösen“ diese Schwelle rund 62 %:

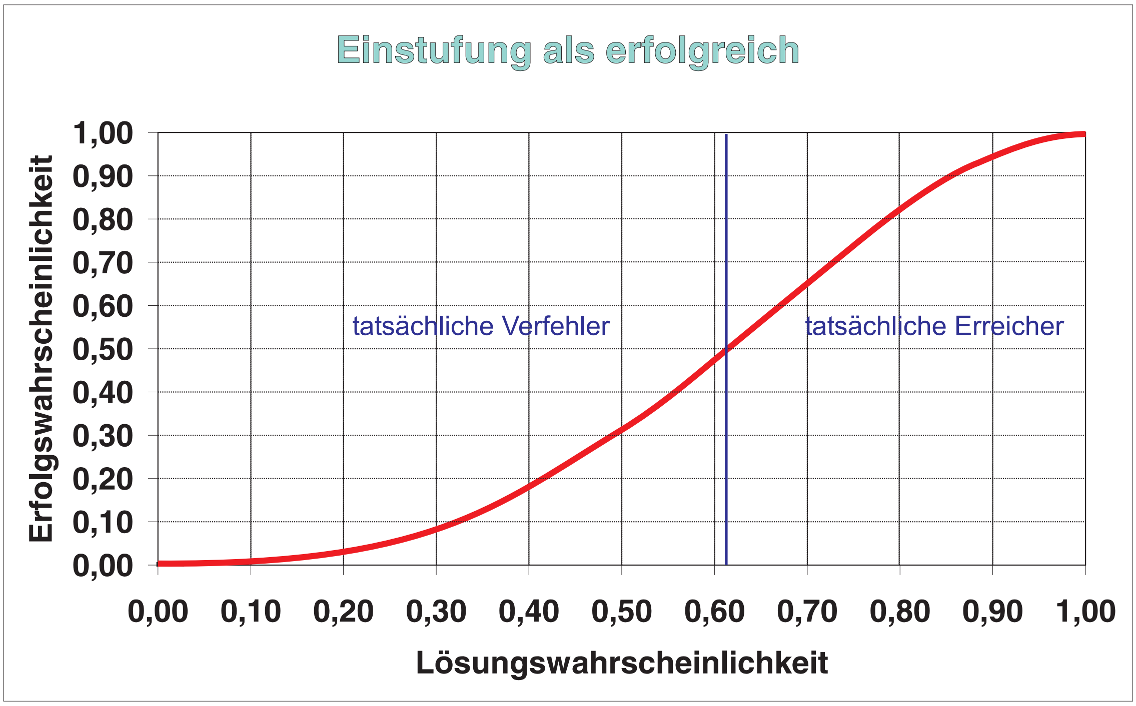


Fig. 7

Bei einer solchen Wahl haben alle Prüflinge höchstens das Fehlbeurteilungsrisiko 50 % und dieses Risiko sinkt mit wachsendem Abstand von p_0.

I.3 Modellierung von Aufgabenschwierigkeit und Schülerfähigkeit

Aufgabencharakteristiken im Raschmodell. Es soll nun eine Modellvorstellung skizziert werden, die bei der Interpretation von PISA-Testaufgaben und IGLU-Testaufgaben verwendet wurde und wird:
Schritt 1: Wahrscheinlichkeiten sind als „Fähigkeitsmaß“ schlecht geeignet, da sie auf das Intervall $[0;1]$ eingeschränkt sind. Es scheint glaubhaft, dass eine Steigerung der Bearbeitungssicherheit von 40 % auf 60 % weniger Lern- und Übungsaufwand erfordern kann als eine Steigerung von 95 % auf 98 %. Daher bieten sich zur Umrechnung von Lösungswahrscheinlichkeiten in „Fähigkeitsmaße“ Funktionen an, die das Intervall $]0;1[$ auf ganz $\mathbb{R}$ abbilden. Eine solche

Funktion ist z.B. der logarithmierte Chancenquotient $\xi(p)$, bei dem das Verhältnis $\frac{p}{1-p}$ der Lösungswahrscheinlichkeit p zur Fehlerwahrscheinlichkeit $1-p$ logarithmiert wird:

$$\xi(p) = \ln \frac{p}{1-p}$$

Gibt man einen Wert ξ vor, so berechnet sich die zugehörige Lösungswahrscheinlichkeit p in der Form:

$$p = \frac{\exp(\xi)}{1+\exp(\xi)}$$

Der Funktionsterm ist als Term der sogenannten *logistischen Funktion* bekannt [3].

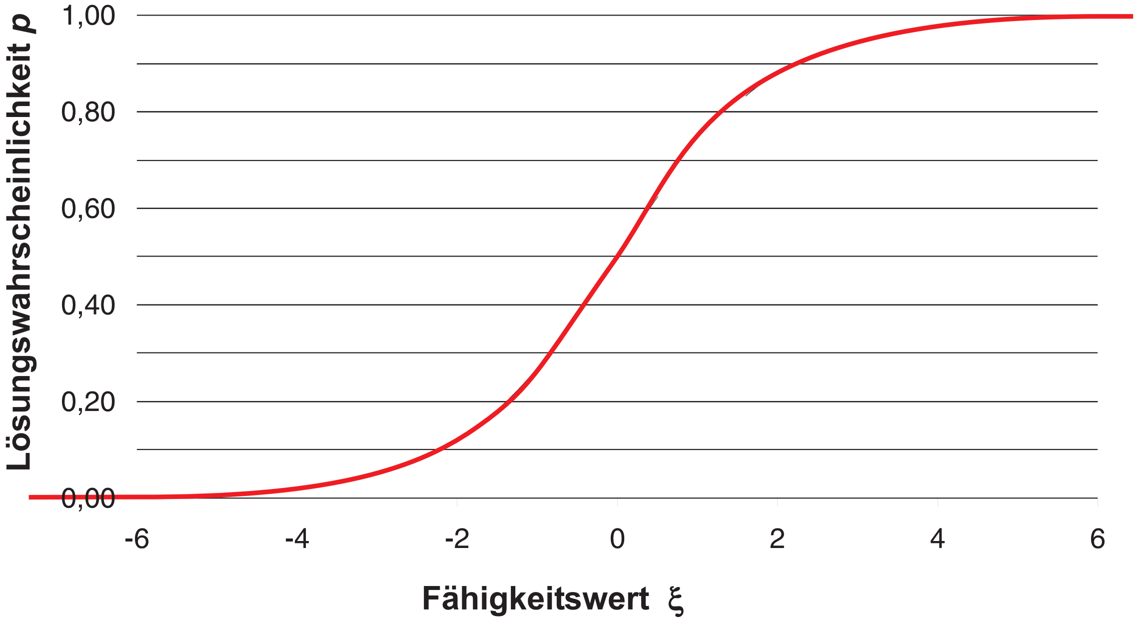


Fig. 8

Unterstellt man nun, dass sich bei verschieden schweren Aufgaben jeder Aufgabe eine hemmende Gegenwirkung σ zuschreiben lässt, die den Wert ξ verkleinert, so ist die einfachste Modellvorstellung das Subtrahieren von σ.

Damit erhält man für einen Probenden mit Fähigkeitswert ξ (der Wert sei über eine „Standardaufgabe“ definiert, der man den *Schwierigkeitswert 0* zuschreibt) bei einer Aufgabe mit Schwierigkeitswert σ die Lösungswahrscheinlichkeit

$$p = \frac{\exp(\xi-\sigma)}{1+\exp(\xi-\sigma)} = \frac{1}{1+\exp(\sigma-\xi)}\,.$$

Man nennt diesen Ansatz *Rasch-Modellierung*.

[3] Diese Funktion tritt bei manchen bechränkten Wachstumsprozessen auf.

Aufgabencharakteristiken im Birnbaummodell. Nimmt man noch einen „Steilheitsparameter“ δ hinzu, so lassen sich sogar Aufgaben modellieren, für die man zunächst kaum Lösungschancen hat, dann aber kurz nach Überschreiten der Schwelle σ sehr sicher ist:

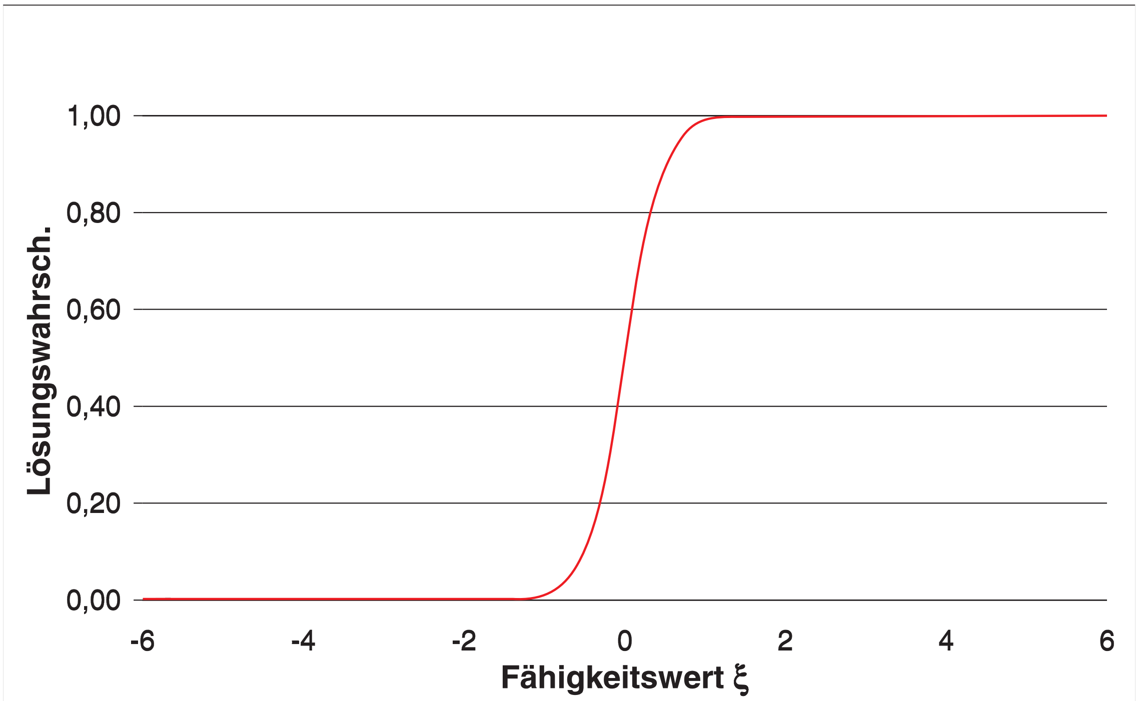


Fig. 9

Hier lautet die Beziehung

$$p = \frac{1}{1 + \exp(\delta \cdot (\sigma - \xi))} .$$

Diese Modellierung wurde 1968 von Birnbaum vorgeschlagen.

Es sind weitere, noch komplexere Modelle „auf dem Markt“, in denen z. B. mehr als eine Fähigkeitsdimension und/oder mehr als nur zwei Bewertungskategorien für die beobachtete Bearbeitungsqualität betrachtet werden. Diese Modelle sind jedoch für den *Evaluationsalltag* in der Schulpraxis noch weniger interessant, als das *Raschmodell* und seine Ableger.

II Schriftliche Prüfung von Lernerfolgen

II.1 Aufgabensystematik nach der Darbietung

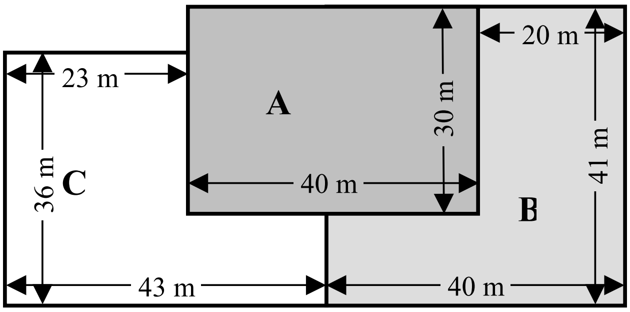
Zum Aufgabenbegriff. Wie bereits ausgeführt wurde, kann man eine Lernerfolgsprüfung als *Darbietung eines Reizmaterials* auffassen, auf das hin eine *Antwort* verlangt wird. Diese beiden Aspekte lassen sich zwar nicht immer genau gegeneinander abgrenzen, ermöglichen jedoch eine Klassifikation nach der *Darbietungsweise,* der *Art und Form* des Materials und der *Art und Weise, wie die Antwort verlangt wird*.
Wir beschränken uns im Folgenden auf Materialien, in denen *Informationen* durch Texte, und grafische Darstellungen (in Zukunft vielleicht sogar Vidosequenzen) gegeben werden. Da es um schriftliche Prüfungen gehen soll, wird(werden) die Antwort(en) in „Schriftform“ verlangt. Damit soll unter *Aufgabe* ein „Informationsobjekt“ verstanden werden, das zusammen mit Aufforderungen zu speziellen Handlungen dargeboten wird.

Einige Darbietungsweisen. Aufgaben lassen sich in der Praxis vorwiegend auditiv oder visuell darbieten, wobei die visuelle Darbietung (Informationsaufnahme durch Lesen von Texten, Interpretation von Bilder und Grafiken oder das Verfolgen von Videoclips, . . .) überwiegt.
Es kommt dabei sowohl auf den intendierten Inhalt der Aufgabe als auch die „technischen“ Möglichkeiten des Aufgabenstellers und seiner Adressaten (sie sollen wie in der Testpsychologie üblich *Probanden* genannt werden) an.
Als erstes Beispiel soll die Aufgabe „Grundstücke“ aus den Lernstandserhebungen NRW 2005 für den Jahrgang 9 dienen, in der die Bestimmung von Flächeninhalten in Figuren geprüft wird, die sich als Vereinigung von Rechtecken beschreiben lassen. Hier geht es inhaltlich nicht nur um den Flächeninhalt eines isolierten Rechtecks. Es ist vielmehr beabsichtigt, die Probanden zur Erarbeitung und Anwendung eigenständiger Zerlegungsstrategien zu veranlassen. Damit bot sich für die Aufgabenkonstrukteure eine grafische Darstellung von Flächen mit nicht redundanten Bemaßungen an.
Das Frageformat kann als Koppelung von *freier Antwortnotation (Berechnung)* mit *Ausfüllen von Antwortlücken* gekennzeichnet werden.

Grundstücke

Wie groß sind die rechts abgebildeten Grundstücke A, B und C?

Berechne jeweils den Flächeninhalt!



Platz für Berechnungen:

Ergebnis:

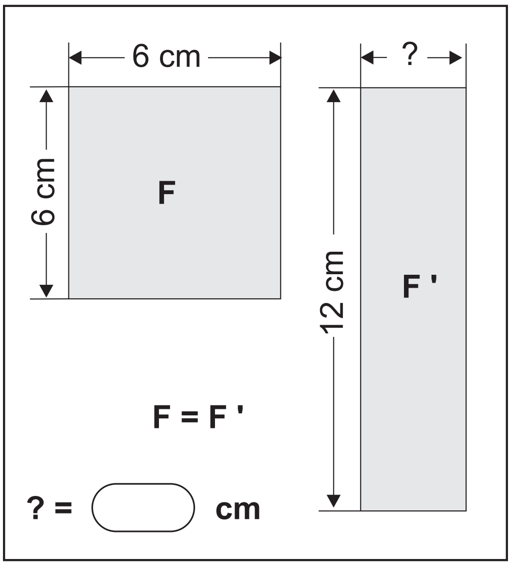
Der Flächeninhalt von Grundstück A beträgt ______ m².

Der Flächeninhalt von Grundstück B beträgt ______ m².

Der Flächeninhalt von Grundstück C beträgt ______ m².

Dabei ist allerdings die Notation von Berechnungen nicht zwingend vorgeschrieben und soll wohl eher der Fehlerdiagnostik zwecks Kategorisierung von Fehlantworten dienen.
Die Aufgabe ist ihrem Format nach eine Kombination von *verbalen* und *nicht-verbalen* visuellen Elementen. Da die Textteile rudimentär sind, würde eine auditive Fassung für leistungsschwache Probanden nur wenig nützen (solche Fassungen setzen auf das Vorlesen von Texten und die Präsentation der zugehörigen Grafiken).
Wie das folgende Beispiel einer *nicht-verbalen* und einer *verbalen* Fassung desselben Aufgabeninhalts zeigt, gibt es auch geometrische Aufgaben in „Reinform“:

nicht verbal



verbal

Ein Quadrat mit der Seitenlänge 6 cm wird in ein flächengleiches Rechteck verwandelt. Eine Seite des Rechtecks ist 12 cm lang.

Wie lang ist die andere Seite?

- ☐ 1 cm
- ☐ 2 cm
- ☐ 3 cm
- ☐ 4 cm
- ☐ 5 cm

Fig. 11

Hier kann man natürlich trefflich darüber streiten, ob wirklich dasselbe Lernziel geprüft wird: In der nicht-verbalen Form muss erst einmal die Information der Flächengleichheit aus formelmässigen Notationen entnommen werden. Dann muss sich der Proband noch selbst die Frage und die Aufforderung zum Ausfüllen der Antwortlücke „zusammenreimen“ (wer mit solchen Aufgabenformen vertraut gemacht worden ist, hat diese Schwiergkeiten allerdings nicht in solchem Ausmaß).
Bei der verbalen Form fällt auf, dass durch das Antwortformat (Ankreuzmodus!) eigentlich nur geprüft wird, ob der Proband die Flächeninhaltsformel fünfmal korrekt anwenden kann. Man kann nämlich durch Einsetzen der Alternativen in die Flächeninhaltsformel sofort die richtige Alternative finden. Die nicht-verbale Fassung ist zwar auch so lösbar, verlangt aber dann, dass der Proband sich seine „Probewerte“ selbst wählt.
Dass sich hier aus den beiden Aufgabenformen eine auditive Fassung mit Präsentationsgrafik machen lässt, liegt hier klar auf der Hand. Liest man die Textfassung vor und zeigt eine weniger überladene Fassung der Grafik, so dürfte die Aufgabe Leseschwachen leichter fallen.
Eine typische „Gymnasialfassung alter Prägung“ wäre folgende grafikfreie Fassung:

> *Ein Quadrat mit der Seitenlänge* $a = 6$ *cm wird in ein flächeninhaltsgleiches Rechteck verwandelt, dessen eine Seitenlänge 12 cm beträgt.*
>
> *Bestimme die andere Seitenlänge.*

Für schwache Leser dürfte sich hier die Aufgabenschwierigkeit schon durch den Relativsatz etwas erhöhen.

In der Schulpraxis werden Lernerfolge und Leistungen überwiegend durch Gruppenprüfungen mittels „Tests“ (dazu zählen auch Klassenarbeiten) geprüft. Dann spielt auf jeden Fall wegen der Vorlage von *mehreren* Aufgaben die Begrenzung der Bearbeitungszeit eine Rolle.
Die Testpsychologie unterscheidet dazu *Schnelligkeitstets* von it Niveautests. Die ersteren setzen sich aus vielen kleineren *Items* (=Testaufgaben) zusammen, die für die Adressaten „mittelschwer“ bis „leicht“ sind. Niveautests bestehen dagegen aus relativ wenigen „mittelschweren“ bis hin zu „schwierigen“ Testitems, die vor allem höhere Anforderungen an die Denkfähigkeit der Probanden stellen. Als Beispiel für eine solches Item mag die bekannte „31 Pf - Aufgabe“ aus dem PISA 2000 - Test dienen:

> *Wie kannst du einen Geldbetrag von genau 31 Pfennig hinlegen, wenn du nur 10-Pfennig-, 5-Pfennig- und 2-Pfennig-Münzen zur Verfügnung hast.*
> *Gib* ***alle*** *Möglichkeiten an!*

Bei dieser Aufgabe fanden nur 2,9 % aller geprüften deutschen 15jährigen alle 6 Möglichkeiten. Vier bis fünf Möglichkeiten fanden immerhin 17,9 % der Prüflinge. Der „numerische“ Anspruch dieser Aufgabe ist offensichtlich gering und liegt auf Grundschulniveau. Dies gilt jedoch nicht für den *kognitiven* Anspruch, da die Aufgabe nur mit der Erprobung und Kontrolle von Systematisierungsstrategien lösbar ist[1]

Zum Schluss dieser Darstellung soll noch auf den Unterschied zwischen Einzel- und Gruppentestung und den Einfluss der Aufgabenpositionierung im Test hingewiesen werden:

- Bei der *Einzeltestung* lässt sich zwar die gegenseitige Beeinflussung von Probanden vermeiden, dafür wirkt sich oft der Einfluss des *Prüfers* auf die getestete Person stärker aus (man denke hier nur an den Umgang mit Prüfungsängsten).

- Bei der Gruppentestung ist der unmittelbare Einfluss des Prüfers auf Einzelpersonen geringer. Dafür sind diese der „Grundstimmung“ im Prüfungsraum ausgesetzt oder werden sogar durch „Hilfeersuchen“ von Mitprüflingen gestört.

Bei manchen Aufgaben spielt es durchaus eine Rolle, an welcher Stelle des Tests sie gestellt wird. Dieser sogenannte *Positionseffekt* beruht darauf, dass bestimmte Aufgabenserien beim Bearbeitung eine Einstellung beim Probanden aufbauen können. Die erste daruf folgende Aufgabe, die *nicht* nach dem vorherigen „Schema“ lösbar ist, fällt dann schwerer, da sie eine Neueinstellung des Probanden verlangt.
Werden Parallelformen von Tests durch reine Umordnung der Aufgabenplätze erzeugt, so muss darauf geachtet werden, dass keine der Formen solche Einstellungseffekte fördert. Es ist bekannt, dass bei vorgeschriebener Bearbeitungsreihenfolge die erste Aufgabe eines Tests etwas schwerer fällt als es bei einer Platzierung an späterer Stelle wäre.

[1]Die 6 Zerlegungen sind übrigens $13\times2+1\times5$, $8\times2+3\times5$,$3\times2+5\times5$, $8\times2+1\times5+1\times10$, $3\times2+1\times5+2\times10$, $8\times2+3\times5+1\times10$.

II.2 Aufgabensystematik nach dem Antwortformat

Freie Aufgabenbeantwortung. Im Gegensatz zu gebundenen Antwortformen kann der Proband ohne besondere Einschränkungen seine Antworten auf Fragen selber formulieren. Wir setzen im Folgenden voraus, dass eine Aufgabe aus einem oder mehreren Informationsblöcken und zugehörigen Fragen besteht.
Neben der im Mathematikunterricht eher unüblichen Aufforderung *zu freier Äußerung* im Sinne eines *Essays* (z. B. „erzähle eine Geschichte über die Zahl π", „beschreibe, wo das Rechnen mit Brüchen gebraucht wird", ...) finden sich hier viele Varianten von Antwortformaten, die man unter *freie Aufgaben-Bearbeitung mit eingeschränktem Antwort-Spielraum*[2] zusammenfassen kann:

Kurzantwort-Aufgaben
Hier werden Fragen gestellt, die mit einem Wort, einer Zahl oder einem Satz beantwortet werden sollen.

Beispiel 1:
Ist $\frac{2}{3}$ kleiner als $\frac{5}{7}$? ______________________

Beispiel 2:
Wie lautet der erste Strahlensatz? ____________________________________

Assoziations-Aufgaben
Hier geht es um das Erlernen von Paaren. Ein Paarling wird vorgegeben, der andere ist vom Prüfling zu nennen.

Beispiel 1:
Nenne den jeweils den Kehrbruch:

$\frac{7}{8}$ _____ $\frac{9}{5}$ _____ $\frac{2}{3}$ _____ $\frac{1}{6}$ _____

Beispiel 2:
Bestimme jeweils den Umfang des Kreises mit Radius r:

$r = 2{,}2$ cm ________ $r = 3{,}5$ cm ________ $r = 1{,}5$ cm ________

Beide Beispiele sind eigentlich keine echten Assoziationsaufgaben im Sinne der Testpsychologie, da der zweite Paarling weniger *memoriert* als *bestimmt* werden muss. Am ehesten kann ein Memorieren noch in folgendem Fall erwartet werden:

Beispiel 3:
Nenne den jeweils die Gegenzahl:

3 _____ -4 _____ 0 _____ $2{,}5$ _____

[2] vgl. F. Beiner (1982), Prüfungsdidaktik und Prüfungspsychologie, S. 169 ff

Ergänzungs-Aufgaben
Die *Antwortschablonen* sind Lückentexte, unvollständige Tabellen oder Grafiken, in die etwas so einzutragen ist, dass sich ein sinnvoller Zusammenhang ergibt.

Beispiel 1: vgl. die Aufgabe „Grundstücke" aus den Lernstandserhebungen NRW.

Beispiel 2:
Wenn man den Kreisradius verdoppelt, wächst der Kreisumfang aus das _____ fache und der Kreisflächeninhalt auf das _____ fache.

Mathematische Lang-Antwort-Aufgaben
Hierher gehören die typischen Aufgabenformate der Klassenarbeiten höherer Jahrgangsstufen, in denen Arbeitsanweisungen zu befolgen und *Rechenwege* oder gar *Argumentationsketten* zu dokumentieren sind. Wir geben wegen der Komplexität solcher Aufgaben nur ein (anspruchsvolles!) Beispiel aus der Geometrie und verweisen ansonsten auf die in der ersten Übungsserie angesprochene Aufgabe „Traktor" aus den Lernstandserhebungen NRW.

Beispiel:
Eine kegelförmige Filtertüte sei mit einer Flüssigkeit vollständig gefüllt, welche durch ein Loch in der Spitze auslaufen kann. Die Höhe des Kegels sei dreimal so groß wie der Radius r seines Grundkreises.

(1) Wie hoch steht die Flüssigkeit noch, wenn der Durchmesser des Flüssigkeitsspiegels auf die Hälfte gesunken ist?

(2) Wie hoch steht die Flüssigkeit noch, wenn das Volumen der Flüssigkeit auf die Hälfte gesunken ist?

Zum Einsatz und zur Gestaltung von Ergänzungs-, Kurzantwort- und Assoziationsaufgaben kann man folgende Empfehlungen geben, die bereits R.L. EBEL 1951 formuliert hat. Die fünfte Empfehlung ist zwar im Zeitalter des Taschenrechners im zweiten Teil nicht mehr aktuell bzw. sogar kontraprodutiv, soll aber mit aufgeführt werden.

1. Benutzen Sie Kurzantwort-Formen nur für Aufgaben, die durch ein einziges Wort bzw. einen einzigen Satz bzw. eine einzige Zahl gelöst werden können.

2. Übernehmen Sie den (verwendeten) Satz nicht wörtlich aus einem zusammenhängenden Text, um ihn ergänzen zu lassen bzw. in eine Frage umzuformen.

3. Formulieren Sie die Frage bzw. den zu ergänzenden Satz im Hinblick auf eine exakt definierte Antwort. (Gegen diese Regel verstößt z. B. die Frage: „Wer war Bismarck?")

4. Lassen Sie ausreichend viel Platz für die Antwort und die Bewertung frei.

5. Geben Sie bei Rechenaufgaben den gewünschten Genauigkeitsgrad an. *Am besten sollte man die Aufgaben so einrichten, dass nur ganze Zahlen herauskommen, sofern nicht der Umgang mit Brüchen und Dezimalzahlen geprüft werden soll.*

6. Bei Ergänzungsaufgaben dürfen die Lücken nicht so dicht aufeinander folgen, dass der Sinn des Textes verloren geht.

Gebundene Aufgabenbearbeitung. Bei der gebundenen Aufgabenbeantwortung sind verschiedene Antwortmöglichkeiten vorgegeben. Der Prüfling muss davon eine oder auch mehrere als richtig oder falsch kennzeichnen oder bestimmte Zuordnungen vornehmen.

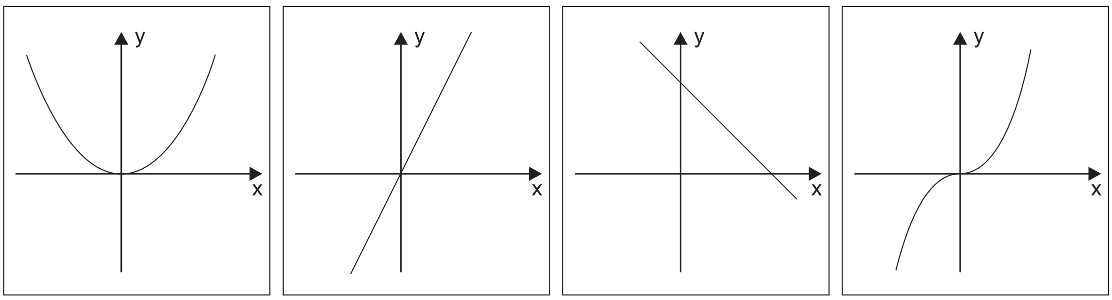
Umordnungs- und Zuordnungsaufgaben
Umordnungs-Aufgaben verlangen vom Adressaten, eine beliebige Folge von Aussagen oder Symbolen in eine sinnvolle Reihenfolge zu bringen.

Beispiel 1:
Ordne die rationalen Zahlen der Größe nach, fange mit der kleinsten an: 2; $-0{,}5$; $\frac{2}{3}$; $\frac{3}{5}$; -2.

Beispiel 2:
Zu welchem der Graphen passt welche Funktionsgleichung? Trage die Nummer unter dem Graph ein.

Funktionsgleichungen:
(1) $y = x^2$ (2) $y = 2x$ (3) $y = x^3$ (4)$y = 4 - x$

Graphen:



( ) ( ) ( ) ( )

Fig. 12

Auswahlantwort-Aufgaben

Zweifachwahl-Aufgaben (single-choice)
Dieser Aufgabentyp bietet nur zwei Alternativen als mögliche Antworten an, z. B. „Ja/Nein“, „Richtig/Falsch“, „Größer/Kleiner“, ...

Beispiel 1:
Wenn ein Kapital ohne es anzurühren zwei Jahre lang mit jeweils mit dem Zinssatz 3 % verzinst wird, wächst es insgesamt um 6 %.

Stimmt diese Behauptung? ☐ ja ☐ nein

Beispiel 2:
Setze das richtige der Zeichen $<$ und $>$ zwischen die Brüche:

a) $\frac{7}{8} \quad \frac{6}{7}$ b) $\frac{25}{13} \quad \frac{9}{5}$ c) $\frac{4}{7} \quad \frac{3}{5}$

Der Nachteil dieses Aufgabentyps ist, dass die Ratewahrscheinlichkeit für die richtige Alternative 50 % beträgt, wenn jemand „blind rät“. Dies lässt sich nicht vermeiden und macht solche Aufgaben in isolierter Form für Massenprüfungen fast unbrauchbar. Kombiniert man wie in Beispiel 2 k solche Items und verlangt, dass sie alle gelöst sein müssen, so reduziert sich die Ratewahrscheinlichkeit auf $0,5^k$.

Der Aufgabenkonstrukteur muss darauf achten, dass das Entscheidungskriterium für die Auswahl der Antwort für die Probanden klar ist und nicht etwa mehrere Kriterien denkbar sind. Dies ist insbesondere dann zu befürchten, wenn lange Sätze und schwierige Satzkonstruktionen in der Aufgabenformulierung vorkommen.

Mehrfachwahl-Aufgaben (multiple-choice)
Hier ist üblicherweise unter mehreren Alternativen nur eine richtig und soll ausgewählt werden. Es gibt allerdings auch Items (vgl. die offiziellen Tests im Theorieteil der Fahrprüfung!), bei denen mehr als eine Alternative richtig ist.

Multiple-choice Items haben den Vorteil hoher Auswertungsobjektivität und der Möglichkeit der automatisierten Auswertung. Da dies mit einer deutlich reduzierten Ratechance gekoppelt ist, sind sie in Tests sehr beliebt.

Als erstes Beispiel möge die verbale Form der bereits vorgestellten Rechtecksaufgabe dienen.

Beispiel 2:
Eine Regionalbahn fährt um 16:32 Uhr in Astadt Hbf ab. Sie legt durchschnittlich 90 km pro Stunde. Wann kommt sie in Bstadt Hbf an (Entfernung 110 km)?

☐ um 18:02 Uhr

☐ um 17:45 Uhr

☐ um 18:02 Uhr

☐ um 19:02 Uhr

☐ um 18:45 Uhr

Das letzte Beispiel lässt schon erahnen, dass man die *Fehlantworten* (im Folgenden, wie in der Testpsychologie üblich, *Distraktoren* genannt), nicht „willkürlich" wählt.

Zuerst löst der Konstrukteur die Aufgabe im „freien Modus" selbst, um dann die Distraktoren möglichst so zu wählen, dass mögliche Fehlvorstellungen, übliche Rechenfehler, . . . abgebildet werden. Hier ergibt sich zunächst eine Fahrtzeit von etwa 73,3 Minuten, also fahrplanmäßig 1 Stunde und 13 Minuten. Daraus resultiert die richtige Antwort „um 17:45 Uhr". Der Distraktor „um 18:02 Uhr" ist für einen Schüler attraktiv, der mit 90 Minuten Fahrtzeit arbeitet. Die beiden anderen Alternativen setzen eher an „äußerlichen Rateüberlegungen" an.

Man muss sich allerdings darüber im Klaren sein, dass „attraktiv" scheinende Distraktoren für unwissende Probanden manchmal nicht plausibler als andere Alternativen wirken. Dann kann sich bei einer Aufgabenanalyse in einer großen Population ein Bild wie in Fig. 13 ergeben (es handelt sich um eine MC-PISA-Aufgabe aus dem Bereich *Geometrie*):

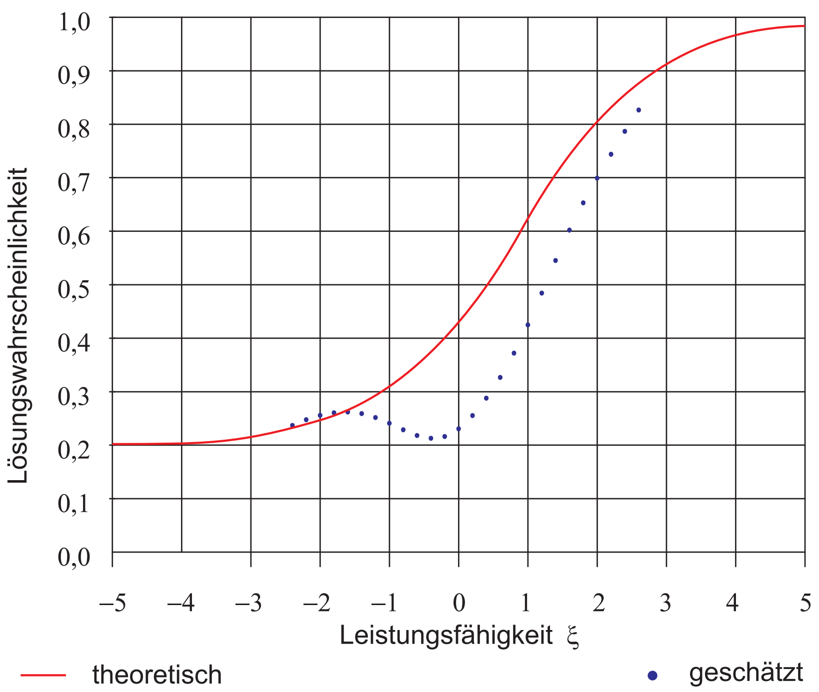


Fig. 13

Offensichtlich raten hier leistungsschwache Probanden „blind“, Probanden mit „etwas zu viel und noch nicht genug Wissen“ fallen auf mindestens einen der Distraktoren herein, und Probanden mit genügend Hintergrundwissen bevorzugen mit wachsendem Können immer mehr die richtige Alternative.

Zur Aufgabenformulierung kann man folgende Ratschläge geben:

Die Aufgabenstellung sollte eine klar formulierte Arbeitsanweisung enthalten, die dem Prüfling sagt, was er tun soll.
Diese Arbeitsanweisung kann folgende Form haben:

– eine Frage,

– einen Arbeitsauftrag (ggf. unter Verwendung von Arbeitsmaterial)

– eine Problemstellung mit der Aufforderung zur Lösung

Die Formulierung der Aufgabe muss möglichst knapp und unmissverständlich sein. Die Leistung des Prüflings darf nicht darin bestehen, dass er eigentlich nur zu deuten oder erraten hat, was der Prüfer verlangt. Auch ein Schüler, der den Stoff nicht beherrscht, muss aus der Aufgabenstellung erkennen können, welche Leistung verlangt wird. Damit sind sachfremde Erschwernisse ebenso unnötig wie überflüssige Lösungshilfen.

Gegenbeispiel zum letzten Ratschlag:

Wie groß ist der Flächeninhalt eines Rechtecks mit den Seitenlängen 4 cm und 6 cm?
☐ 12 cm ☐ 24 cm^2 ☐ 10 cm ☐ 24 cm

Aus der Aufgabenstellung geht hervor, dass nach einem Flächeninhalt gefragt wird. Wer weiß, dass hier Flächeninhalte in cm^2 gemessen werden, schliesst sofort auf die richtige Antwort. Dies ist wohl nicht im Sinne des Prüfers.

Allgemein wird ein solches Verwenden „unlogischer Distraktoren“ die Ratewahrscheinlichkeit nicht vermindern und ist daher sinnlos.

Fragen dürfen nicht zu allgemein formuliert werden und notwendige Annahmen müssen explizit gemacht werden.

Gegenbeispiel:

Marc braucht für seinen 1,5 km langen Schulweg 21 Minuten. Stefanie wohnt 2 km von der Schule entfernt. Wie viele Minuten braucht sie wohl?

☐ 25 ☐ 28 ☐ 31 ☐ 34

Wird die Aufgabe in der Unterrichtsperiode „Dreisatzrechnen“ gestellt, so werden fast alle Schüler gleiches Gehtempo annnehmen, obwohl dies nicht gesagt wurde. Dann kommt nur 28 als richtig in Frage.

Am Ende des 8. Schuljahres könnte das Wort „wohl“ vor dem Fragezeichen allerdings von manchen Prüflingen ernst genommen werden. Dann lassen sich mindestens die benachbarten Alternativen von 28 begründen (z.B. Marc ist nicht der Schnellste und Stefanie kann es durchaus in 25 Minuten schaffen).

Die Distraktoren sind von großer Bedeutung für die Schwierigkeit und die Güte einer Testaufgabe. Der Lerner soll nur mit Sachverstand richtigen und falschen Antworten unterscheiden können. Alle angebotenen Distraktoren sollten im Zusammenhang mit der Aufgabenstellung sinnvoll, logisch und gleich wahrscheinlich erscheinen.

Über diesen Punkt wurde schon in den vorhergehenden Abschnitten vieles gesagt. Der Ratschlag ist als Appell an den Aufgabenkonstrukteur zu verstehen, sich in Prüflinge mit wenig, mittlerem und viel Wissen hinein zu versetzen und über die Wirkung intendierter Distraktoren nachzudenken. Dies kann ihn zwar vor offensichtlichen Formulierungsmängeln bewahren, wird ihn jedoch nicht immer vor den selbsernannten „Besserwissern“ schützen, die nachträglich „vermeidbare Ratemöglichkeiten“ entdecken!

Eine Hilfe zur Vermeidung grober Fehler dürften auf jeden Fall die Ratschläge von EBEL in dem bereits zitierten Aufsatz sein, die wir sinngemäß aus Seite 97 von KLAUER et al. (1972) übernehmen[3]:

1. Benutzen Sie entweder eine Frage oder einen unvollständigen Satz als „Aufgabenstamm“.
2. Formulieren Sie den Aufgabenstamm so, dass Sie nicht in jeder Antwortalternative ein bestimmtes Wort wiederholen müssen.
3. Falls eine *negative* Auswahl vom Adressaten verlangt wird, so heben Sie dies ganz besonders hervor (z. B. durch Fettdruck oder Unterstreichen). Fassen Sie möglichst die Aufgaben mit negativer Auswahl zu einem Block zusammen.
4. Bemühen Sie sich, die „beste Antwort“ und ihre Alternativen so zu formulieren, dass sie auch von Fachleuten als „beste Antwort“ akzeptiert wird.
5. Alle Alternativen müssen grammatikalisch auf den Aufgabenstamm abgestimmt sein.
6. Gestalten Sie die Distraktoren einsichtig und attraktiv für jene Adressaten, die nicht über die abgefragten Fähigkeiten bzw. das abgefragte Wissen verfügen.
7. Vermeiden Sie Distraktoren, die wesentliche über oder unter dem angenommenen Niveau der Adressaten liegen. Sie werden in der Regel ohnehin nicht gewählt.
8. Vermeiden Sie Antworten, die sich gegenseitig voraussetzen oder einschließen.
9. Benutzen Sie die Alternative „keine von diesen“ nur bei Aufgaben, für die genau ein richtiges Ergebnis angegeben werden kann.
10. Bringen Sie die Antworten möglichst in eine sinnvolle Reihenfolge. Vermeiden Sie aber, dass die richtige Antwort immer am selben Platz steht.
11. Wird die Definition eines Begriffes abgefragt, so ist es günstiger, den Begriff in den Aufgabenstamm zu nehmen und alternative Definitionen anzubieten.

[3]Klauer, K.J., Fricke, R., Rupprecht, H., Schott, F.: Lehrzielorientierte Tests. Düsseldorf 1972

II.3 Lehrzielorientierte informelle Tests

Zielsetzung für schriftliche Prüfungen. Von einem schriftlichen Prüf- und Messverfahren ist zu verlangen, dass es eine relativ hohe „Testgüte“ (darüber später mehr) ermöglicht, spezielle Lehrziele von Unterricht und Ausbildung abdecken kann und vom Lehrer/Dozenten/Ausbilder selbst erstellt werden kann. Zudem soll es sich im Laufe der Berufspraxis optimieren und vervollständigen lassen.

Falls um die Prüfung von Leistungsbereichen geht, die sich für schriftliche Erfolgskontrollen eignen, bieten sich sogenannte *informelle Tests* an, die an den angestrebten Lehrzielen orientiert sind. Solche Tests lassen sich nämlich sowohl unter *didaktischen* und *curricularen* als auch unter *messtheoretischen* Gesichtspunkten optimieren und langfristig auch ökonomisieren.

Der *Leistungsbegriff* ist dabei insofern in einem umfassenden Sinn zu verstehen, als sich sowohl die mehrkategorielle *Zensurengebung* als auch die dichotome Beurteilung von Lernerfolgen über die gleichen Hintergrundmodelle deuten lassen.

Entwicklungsschritte. Die Entwicklung *lehrzielorientierter informeller Tests* (L-I-T) kann man in drei Abschnitte mit insgesamt neun Schritten unterteilen:

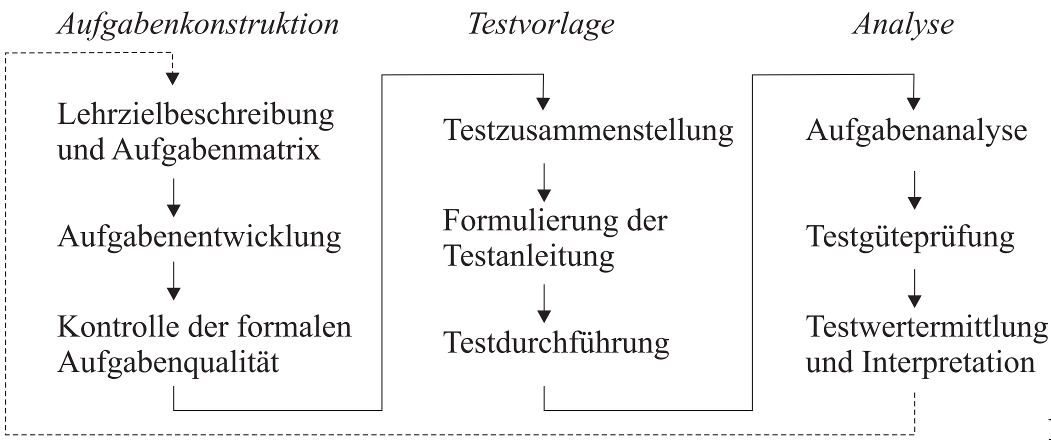


Fig. 14

Die oben aufgeführten neun Schritte sollen nun ihrem natürlichen Ablauf entsprechend beschrieben werden. Nach Durcharbeitung des folgenden Abschnitts sollte jede Lehrperson einen solchen L-I-T erstellen, durchführen und auswerten können.

Die bei der Testanalyse gewonnenen Erfahrungen sollten bei der Neuerstellung oder Veränderung eines Tests berücksichtigt werden. Insofern schließen sich die Entwicklungsschritte zu einem Kreis. Dies soll die gestrichelte Linie andeuten.

Erstellung von Lehrzieltests.

Stoffbereiche und Aufgabenmatrix

Lehrzielorientierte informelle Tests sollen den Lehr-Lern-Erfolg eines Unterrichtsabschnitts überprüfen. Über Erfolg oder Misserfolg kann aber nur anhand eines vorher festgelegten Kriteriums entschieden werden.

Werden Lehrziele wie in Abschnitt I.2 operationalisiert und mit (im Idealfall aus Hintergrundvorstellungen abgeleiteten) Kriterien für Beurteilung des Lernerfolgs versehen, so sollen sie im

Folgenden *Leistungsziele* genannt werden. Nur bei solchen Zielen kann der Grad der Zielerreichung geprüft werden.

Auf jeden Fall sollte bei der *erstmaligen* Konzeption eines umfassenden Tests (oder der erstmaligen Zusammenstellung einer Klassenarbeit) die Grobstruktur des Unterrichtsabschnitts in einer *Lehrzielmatrix* erfasst werden. Diese kategorisiert die angestrebten Lehrziele üblicherweise nach *Stoffbereichen* und *Anforderungsebenen*.

Liegen die Unterrichtsziele in operationalisierter Form vor, so legt jede Zelle der Matrix eine Aufgabenklasse fest, aus der man Testaufgaben entnehmen kann. Fehlt die Operationalisierung, so ist etwas mehr Aufwand nötig und man tut gut daran, in diesem Fall selbst überlegte Aufgabenbeispiele zu einem Unterrichtsziel von Fachkollegen auf ihre „Repräsentativität" hin beurteilen zu lassen.

In jedem Fall resultiert aus dem Vorgehen eine *Aufgabenmatrix,* in deren Zellen sich die vorgesehenen Aufgaben finden. Für ihre Aufstellung empfiehlt sich der folgende Weg.

Festlegung der Stoffbereiche

Zunächst muss der Lehrer auflisten, welche Stoffbereiche im zu überprüfenden Unterrichtsabschnitt erarbeitet wurden. Jedem dieser Stoffbereiche wird dann eine Gewichtungszahl zugeordnet, die ihre Bedeutung innerhalb des gesamten Unterrichtsabschnitts zum Ausdruck bringt.

Ein brauchbarer Maßstab für eine Gewichtung der einzelnen Stoffbereiche dürfte ihre *Erarbeitungszeit* im Unterricht sein. Dabei ist allerdings vorausgesetzt, dass die Erarbeitungszeit eines Stoffbereichs seiner fachlichen Bedeutung entsprach. Durch eine problematische Unterrichtsmethodik verursachte „Längen" sollten auf keinen Fall in Gewichtungszahlen eingehen.

Wenn einzelne Stoffbereiche für nachfolgende Unterrichtsabschnitte besondere Bedeutung haben, sollte man dies entsprechend berücksichtigen.

Beispiel:

Aufgliederung und Gewichtung der Stoffbereiche einer Themeneinheit aus dem 9. Schuljahr.

Thema: Näherungsrechnen

Stoffbereich	*Gewichtung in %*
Der Fehlerbegriff beim Runden	*20*
Runden von Summen und Differenzen	*40*
Runden von Produkten und Quotienten	*40*

Die Gewichtung der Stoffbereiche bewirkt, dass keiner von ihnen im Test zu kurz kommt. Auf der anderen Seite heißt dies auch, dass keiner zu intensiv geprüft werden soll, weil sich für ihn besonders leicht Aufgaben erstellen lassen.

Entsprechend der Gewichtung sollten den Stoffbereichen dann Aufgaben zugeordnet werden. Im vorliegenden Beispiel könnte man z.B. 20 Einzelaufgaben erstellen, die dann entsprechend der Gewichtung auf die Teilgebiete aufzuteilen wären.

Bei der Festlegung der Aufgabenanzahl muss man Kompromisse eingehen. Aus Gründen der Praktikabilität und Ökonomie strebt man so wenig wie möglich, aus Gründen der Zuverlässigkeit der Testung aber so viel Aufgaben wie möglich an, weil dadurch Zufälligkeiten und Irrtümer reduziert werden können. Da die notwendige Aufgabenzahl vom Aufgabentyp abhängt, gehen wir darauf später noch einmal ein.

Festlegung der Anforderungsniveaus

Als nächstes muss nun geklärt werden, welche Anspruchsniveaus mit den Aufgaben abgeprüft werden sollen. Soll nur *Faktenwissen* oder auch *Anwendungs-*, *Übertragungswissen* und vielleicht auch *Problemlösungsfähigkeit* getestet werden? Zur Unterscheidung verschiedener Leistungsniveaus könnte man die sogenannte Taxonomie von Lernzielen für den kognitiven Bereich von BLOOM u.a. heranziehen, wonach Ziele/Leistungen nach folgenden Ebenen zu unterscheiden sind: *Wissen, Verstehen, Anwenden, Analyse, Synthese, Bewertung*. Aus Praktikabilitätsgründen wird allerdings meistens nur mit drei oder vier, statt mit diesen sechs Kategorien gearbeitet. Dabei werden dann einige BLOOMsche Kategorien zusammengefasst.
Von Testkritikern ist immer wieder zu hören, dass man mit der bei Tests sehr gängigen Auswahl-Antwort-Form kaum mehr als Faktenwissen prüfen kann. Schauen wir uns dazu ein Beispiel an:

Testaufgabe:
Gegeben ist die Funktion mit der Funktionsgleichung $y = 4x - 2$. *Wenn man einen x-Wert um 1 vergrößert,*

(A) *wächst der y-Wert um 2*

(B) *wächst der y-Wert um 4*

(C) *fällt der y-Wert um 2*

(D) *fällt der y-Wert um 4*

(E) *ändert sich der y-Wert mal mehr, mal weniger*

In diesem Beispiel soll der Lerner sein Verständnis der linearen Funktion durch Beurteilung von Behauptungen über das Wachstumsverhalten nachweisen. Dabei wird mehr als nur eine reproduktive Gedächtnisleistung verlangt (falls dieses Problem nicht schon in gleicher oder ähnlicher Form im Unterricht vermittelt wurde). Anhand des Beispiels können wir mit SEELIG feststellen: „Antwort-Auswahl-Aufgaben können nicht nur das Erinnern prüfen, sondern auch die Anwendung von Gedächtnismaterial und auch das Verstehen. Darüber hinaus können solche Aufgaben auch Anforderungen an das Denken stellen“ (G.F. SEELIG, 1970).
Höhere Anwendungs-, Analyse- und Bewertungsleistungen lassen sich also auch mit (testtheoretisch günstigen) geschlossenen Aufgaben prüfen. Man muss „nur“ geeignete Aufgaben finden.
Als nächstes stellt sich die Frage, ob ein Test mehrere, und wenn ja, welche *Schwierigkeitsniveaus* abdecken sollte. Auch hier stehen sich wieder Forderungen der Messtheorie und der didaktischen Praxis gegenüber:
Die sicherste Testaussage erhält man nach „Lehrmeinung“ aus einem sehr homogenen Test, der möglichst nur ein angepasstes Schwierigkeitsniveau abprüfen sollte.
In der didaktischen Praxis aber will man häufig gerade mit Hilfe von Prüfungen und Tests Niveau-Unterschiede der Prüflinge ermitteln. Als Kompromiss wird meist vorgeschlagen, Tests mit 2 bis 3 Anforderungsniveaus oder Intensitätsstufen zu erstellen, wobei jede Stufe noch mit genügend vielen Aufgaben abgedeckt sein muss.

Im nächsten Schritt können wir also eine differenzierte Aufteilung der Aufgaben für unseren Test vornehmen und zwar nach Stoffgebieten und Anforderungsstufen.

Aufgaben-Matrix

Die Aufgaben-Matrix soll die *Aufgabenverteilung im Lehrzieltest* festlegen.

Im bereits vorbereiteten Beispiel:

Thema: Näherungsrechnen

Matrix zur Aufgabenverteilung

	Anforderungsebenen			
Stoffbereich	*Erinnern*	*Anwenden*	*Bewerten*	*Summe*
Der Fehlerbegriff beim Runden	*1*	*2*	*1*	*4*
Runden von Summen und Differenzen	*2*	*6*	*0*	*8*
Runden von Produkten und Quotienten	*2*	*6*	*0*	*8*
	5	*14*	*1*	*20*

Die Gewichtungszahlen der Planungstafel (s. oben) sind hier umgesetzt in Anzahlen von zu erstellenden Aufgaben je Stoffbereich und Anforderungsniveau.
Im Vergleich zu klassischen Klassenarbeiten wirkt die Aufgabenzahl 20 recht hoch. Stellt man jedoch in Rechnung, dass diese 20 Aufgaben eher den Charakter klassischer Teilaufgaben haben, wirkt die Anzahl durchaus realistisch.
Zusammenfassend kann man sagen, daß vor der Erstellung von Testaufgaben folgende Planungsschritte erledigt werden müssen:

1. Festlegung der zu prüfenden Stoffbereiche

2. Gewichtung der Stoffbereiche (Hierbei sollte möglichst schon die vorgesehene Aufgabenzahl berücksichtigt werden.

3. Festlegung der Anforderungsstufen

4. Erstellung der Aufgaben-Matrix für die Aufgabenverteilung. Liegen die Lehrziele schon in operationalisierter Form vor, so kann eine solche Aufgaben-Matrix zur Einordnung und Überprüfung der Lehrziele herangezogen werden. Man kann sie dann als *Lehrzielmatrix* ansehen.

Aufgabenentwicklung

Dieser Schritt ist der aufwendigste und wichtigste bei der Testentwicklung. In Abschnitt II.2 wurden mögliche Aufgabenformen mit ihren Vor- und Nachteilen dargestellt. Außerdem wurden dort Hinweise zur günstigen Gestaltung und Formulierung verschiedener Aufgabentypen gegeben, die man unbedingt beachten sollte.

Wegen ihrer hohen „Objektivität" und geringen Ratewahrscheinlichkeit bei ausreichend vielen Aufgaben je Lehrziel werden in Lehrzieltests für größere Populationen überwiegend MC-Aufgaben verwendet. Will man aus fachlichen und Validitätsgründen auch andere, z.B. halboffene Aufgaben einsetzen, sollte man sich auf zwei Typen je Test beschränken, um die Prüflinge nicht durch Formalia zusätzlich zu verunsichern.

Die Konstruktion geeigneter Distraktoren erfordert didaktische Phantasie und Sachverstand. Sich an die falschen Antworten von Adressaten bei vergleichbaren Aufgaben zu erinnern, ist eine Möglichkeit, Distraktoren zu finden. Eine andere ist: Man stellt die Aufgaben zunächst probeweise ohne Antwortangebot (z. B. in einer anderen Lerngruppe oder im voraufgehenden Ausbildungsjahr). Aus den Fehlantworten werden die als Distraktoren ausgewählt, die häufiger auftreten und eine nachvollziehbare Fehlleistung repräsentieren. Damit gewinnt eventuell sogar etwas an diagnostischer Qualität.

Nach Erstellung aller Aufgaben sollte

- die Eindeutigkeit der Aufgabenstellung,
- die Richtigkeit der Lösungen und
- die Eignung der Distraktoren

im Einzelnen überprüft werden. Das Hinzuziehen eines Kollegen ist dazu oft sehr aufschlussreich. Auch sollte nochmals verglichen werden, ob die in der Aufgabenmatrix festgehaltenen Verteilungen in Gewichtung und Anspruchsniveau der Aufgaben hinreichend eingehalten sind.

Als Beispiel soll wieder die Erfolgsprüfung der Themeneinheit *Näherungsrechnen* dienen. Wir stellen dazu einen unvollständigen Test vor, den die Leser zur Übung unter Beachtung der bereits vorgestellten Aufgabenverteilungsmatrix vervollständigen sollen (die Teilaufgaben werden als Items gezählt!):

Test zum Näherungsrechnen

Aufgabe 1:
Eine Zahl soll auf eine vorgeschriebene Stelle gerundet werden. Die Regel lautet:

Schaue die __________ *Stelle an.*

Steht dort ____________________ *, so wird abgerundet.*

Steht dort ____________________ *, so wird aufgerundet.*

Aufgabe 2: *Wie groß ist der Fehler?*
a) Der Messwert $1,253$ *m wird auf volle cm gerundet.*

Fehler: __________ *m*

b) Der Messwert $12,36$ *km wird auf volle 100 m gerundet.*

Fehler: __________ *km*

Aufgabe 3: *Auf wie viele Nachkommastellen ist hier zu runden?*

a) $1,2 + 20,56$ *auf* ___ *b)* $23,57 \cdot 1,25$ *auf* ___

Aufgabe 4: *Rechne und runde entsprechend.*

a) $1,23 + 12,5 =$ ____________________ *b)* $103,4 - 98,55$ ____________________

Aufgabe 5: *Tina meint, dass die Angaben „1,3 km“ und „1,30“ km auf Wegweisern an Wanderwegen dasselbe bedeuten. Angela widerspricht ihr.*

Kreuze an und begründe deine Meinung:

☐ *Tina hat recht.* ☐ *Angela hat recht.*

Begründung:

Die Aufgaben im letzten Beispiel sind bewusst „verbesserungsfähig“ gehalten, um den Sinn der bisher zitierten Ratschläge zur Aufgabenkonstruktion noch einmal deutlich zu machen. So ist z. B. in Aufgabe 1 zu fragen, ob denn die Erinnerung an die Rundungsregel hier adäquat mit dem Lückentext geprüft wird.
Wenn die Aufgabe von einer Lehrperson durch Löschen von Wörtern in einem „Lehrtext“ hergestellt wurde, wird unter Umständen nur die reine Erinnerung an „Schreibfiguren“ geprüft und es werden Schüler verunsichert, die sich nicht an eine *sinngemäße* Ergänzung des Textes heranwagen. Hier wäre also durchaus über eine Aufgabenvariante nachzudenken, in der zwischen mehreren Regeln auszuwählen ist.
In Aufgabe 2 ist die Vorsilbe „Mess“ im Wort *Messwert* kontraproduktiv, da „Messwerte“ üblicherweise schon als gerundet gelten. Hier werden wohl Schüler begünstigt, die nicht viele Skrupel beim Lesen von Aufgabentexten haben (ist das „gut“???) und die Aufgabe so verstehen, wie sie offensichtlich gemeint ist. Schüler, die sich zu sehr an Schreibkonventionen bei Maßangaben erinnern, verlieren dagegen Zeit beim Nachdenken über den Sinn der Vorsilbe.
Aufgabe 3 enthält eine unnötige Falle, da auch bei der Multiplikation nach der sinvollen Anzahl von *Nachkommastellen* gefragt wird. Da der als ungenauer geltende Wert nur dreiziffrig ist, muss das Ergebnis auf insgesamt drei Ziffern gerundet werden. Es bleibt also nur eine Nachkommastelle, da das Ergebnis zwei Stellen vor dem Komma hat!
Aufgabe 4 entspricht den üblichen Standards, da eine vorab trainierte Technik anzuwenden ist.
In Aufgabe 5 kann jede der beiden Ankreuzvarianten hinreichend begründet werden und muss in einem solchen Fall als richtig gewertet werden. Die Standardantwort „Angela hat recht, weil der Wert 1,30 km viel genauer als 1,3 km ist“ ist ebenso stichhaltig, wie eine geeignete Interpretation der Redewendung „dasselbe bedeuten“ unter realen Wanderbedingungen und eine daraus abgeleitete Zustimmung für Tina.

Maßnahmen zur Testgütesicherung

Ein sorgfältig erstellter L-I-T sollte die zu beurteilenden Merkmale besser erfassen als herkömmliche Prüfverfahren. Er liefert in der Regel zwar nur Aussagen über einen Teil der zu beurteilenden Merkmale, dafür aber *verlässliche Daten*. Bei seiner Erstellung hält man sich nämlich soweit wie möglich an die üblichen Hauptgütekriterien für Tests: *Gültigkeit (Validität), Zuverlässigkeit (Reliabilität)* und *Objektivität.*

Zur Gültigkeit

Was ist *Gültigkeit*? Ein Test ist gültig, wenn er wirklich das misst, was er zu messen vorgibt. Genaugenommen ist damit die positive Beantwortung der folgenden drei Teilfragen erforderlich:

$$\text{Prüft der Test} \quad \left\{ \begin{array}{l} \text{1. überhaupt} \\ \text{2. auch alle} \\ \text{3. auch nur} \end{array} \right\} \quad \text{die Verhaltensmerkmale,}$$

die er prüfen soll?

Die inhaltliche Gültigkeitskontrolle zur Verfügung gestellter Tests muss vom Benutzer (z.B. Prüfer) selbst vorgenommen werden, da eine Übereinstimmung von tatsächlichem Unterrichtsstoff und Testaufgaben nicht garantiert werden kann. Die Frage sollte lauten: „Sind die Testaufgaben wirklich eine repräsentative Stichprobe der Aufgaben, die überhaupt zu meinen

Prüfungszielen formuliert werden können?“ (Curriculare Validität)

Die folgenden Punkte können die Gültigkeit von L-I-Ts bei ihrer Erstellung sichern:

– präzise Lehrzielbeschreibung (Operationalisierung)

– Erstellung einer Lehrziel- oder Aufgabenmatrix *vor* der Aufgabenentwicklung

– klare, unmissverständliche Aufgabenformulierungen

– Beschränkung der Lösungstätigkeit auf die zu messende Leistung (z.B. Entlastung von unnötiger Schreibarbeit)

– hinreichend Zeit für die Aufgabenbearbeitung, wenn die Schnelligkeit nicht gemessen werden soll.

Ein weiters Prüfungsverfahren für die inhaltliche Gültigkeit von L-I-Ts ist das „Expertenrating“. Dabei lässt man voneinander unabhängige „Unterrichtsfachleute“ die inhaltliche Gültigkeit der Aufgaben für den zu messenden Leistungsbereich prüfen und begutachten. Neben einer *qualitativen Bewertung* können zusätzlich *quantitative* Verfahren herangezogen werden, wie z.B. die folgende leicht durchzuführbare Überprüfung:

Man lässt jede Testaufgabe den einzelnen Zellen der Lehrziel- oder Aufgabenmatrix zuordnen, die Ausgangspunkt der Aufgabenerstellung war. Dabei sollten mindestens drei unabhängige Fachleute befragt werden. Kommen die Experten zu gleichen Zuordnungen, so kann der Test als *inhaltlich valide* gelten.

Beispiel:

Ein Teiltest zu zwei Stoffbereichen A und B mit insgesamt 10 Items ist von drei Experten beurteilt worden. Das Ergebnis liegt als „Beurteilungsmatrix“ vor (bei jeder Zelle ist angegeben, wie viele Experten das jeweilige Item dieser Zelle zuordneten):

Stoff-gebiet	Niveau / Item	1	2	3	4	5	6	7	8	9	10
A	Erinnern	3	1								
	Anwenden		2	3							
	Bewerten				3						
B	Erinnern					3			1		
	Anwenden						3	3	1	3	
	Bewerten								1		3

Fig. 15

Angenommen, die Aufgabenverteilungsmatrix bei der Testkonstruktion tatsächlich die ersten vier Items zu Stoffgebiet A und die übrigen zu Stoffgebiet B vor. Wenn dann noch im Stoffgebiet A das Erinnern durch *ein* Item, das Anwenden durch *zwei* Items und das Bewerten durch *ein* Item geprüft werden sollten und im Stoffgebiet B die entsprechenden Itemzahlen 1, 4 und 1 waren, kann der Test schon einmal als überwiegend valide gelten, da 8 der 10 Items übereinstimmend zugeordnet wurden.

Bei den Items 2 und 8 würde man allerdings eine Revision in Erwägung ziehen, um sie besser an ihre Zielsetzung anzupassen.

Die Brauchbarkeit der vorgestellten *Validierungsmethode* steht und fällt natürlich mit der Sachkenntnis und dem „Einfühlungsvermögen“ der befragten Experten. Wenn diese z. B. zu wenig über den vorhergehenden Unterricht wissen, werden sie Aufgaben nicht immer dem erwarteten Anspruchsniveau zuweisen. Falls ein Kollegenteam an einer Schule zusammenarbeitet, ist dieser Effekt sicher geringer als bei einer „externen Validierung“.

In der *Testpsychologie* gibt es noch weitere *Validitätsdefinitionen*, die jedoch für die Anfertigung von L-I-Ts weniger geeignet sind. Mehrere dieser Definitionen laufen nämlich auf eine Vorlage einer Probeversion des Tests in einer Versuchsstichprobe hinaus, damit die Testresultate mit einem dort anwendbaren weiteren Kriterium vergleichen kann (!).

Zur Zuverlässigkeit

Die *Zuverlässigkeit* gibt die Genauigkeit an, mit der ein bestimmtes Merkmal (z. B. der Grad der Lehrzielerreichung) gemessen wird. Man würde einen Test als zuverlässig ansehen, wenn er

– bei beliebig oft wiederholtem Einsatz

– in derselben Personengruppe

– bei konstanter Leistungsfähigkeit

– unter unveränderten Bedingungen

jeweils gleiche oder zumindest annähernd gleiche Messergebnisse liefern würde. Jeder Testbearbeiter müsste also annähernd die gleichen Punktzahlen erreichen bzw. überwiegend die gleichen Beurteilungen erfahren.

Der Test muss demnach so konstruiert sein, dass beobachtete Leistungen nicht zu stark zufällig schwanken.

Die folgenden Maßnahmen können die Zuverlässigkeit von L-I-Ts bei ihrer Erstellung einigermaßen sichern helfen:

– eindeutige, unmissverständliche Aufgabenformulierungen

– gleiche bzw. gleichwertige Aufgaben für alle Prüflinge (Standardaufgaben)

– gleiche Arbeitsbedingungen für alle Prüflinge(Standardbedingungen)

– gleicher Beurteilungsmaßstab für alle Prüflinge(Standardbewertung)

– möglichst mehrere Aufgaben zu jeder besonderen im Test erhobenen Leistung (instrumentelle Zuverlässigkeit).

Unter *instrumenteller Zuverlässigkeit* wird verstanden: Hypothetische Zufallseinflüsse sind bei der Bewertung so gut einschätzbar, dass man zu überwiegend zu stabilen Urteilen kommt. Deshalb sollten einzelne Lehrziele (wie bereits in Abschnitt I.2 diskutiert) durch mehrere Aufgaben überprüft werden. Offensichtlich ist ein Test um so zuverlässiger, je mehr inhaltlich gleiche (homogene) Aufgaben er je Lehrziel hat.

Zur Objektivität

Ein schriftliches Prüfverfahren gilt als *objektiv*, wenn in den Prüfungsabschnitten *Durchführung*, *Auswertung* und *Beurteilung* alle subjektiven Einflüsse der Prüfer ausgeschaltet sind, also verschiedene Prüfer zu den gleichen Ergebnissen kommen.

Folgende Bedingungen können die Objektivität von L-I-Ts bei ihrer Erstellung sichern helfen.

Zur Durchführungsobjektivität:

– gleiche oder gleichwertige Aufgaben für alle

– schriftliche Anweisungen und Beispiele

– vorherige Festlegung der Hilfsmittel

– vorherige Festlegung von Testdauer und Anfangszeitpunkt

– möglichst geringe „Interaktion“ zwischen Testern und Getesteten.

Gegen den letzten Punkt verstoßen z. B. Lehrkräfte sehr häufig, die das „Helfersyndrom“ nicht zurückdrängen können und stumme Hilfen geben, wenn sie während einer schriftlichen Prüfung Prüflinge Fehler machen sehen.

Zur Auswertungsobjektivität:

– Aufgabenformen, die eine eindeutige Lösung haben

– Katalog von Auswertungsgesichtspunkten

– vorherige Festlegung von Wertungspunktzahlen

– bei offenen Aufgaben: Liste der noch richtigen Antworten

– Ausschaltung von Sympathie/Antipathie.

Zur Interpretationsobjektivität:

– Benotung bzw. Einstufung ausschließlich aufgrund erreichter Punkte

– vorherige Festlegung des Bewertungsschlüssels

– vorherige Festlegung der Konsequenzen, die mit bestimmten Ergebnissen verbunden werden sollen.

Testvorlage.
Hier lassen sich nach Fig. 14 die drei Phasen *Testzusammenstellung*, *Formulierung der Testanleitung* und *Testdurchführung* unterscheiden.

Testzusammenstellung
Nach den Maßnahmen zur Sicherung der Gütekriterien kann die Zusammenstellung des Tests erfolgen. Dabei muss zunächst über die *Reihenfolge der Aufgaben* im Gesamttest entschieden werden und bei den MC-Items muss noch einmal die *Reihenfolge der Distraktoren* kontrolliert werden.

Falls es sich um einen reinen MC-Item-Test handelt, wird man eine *Lösungsschablone* herstellen und bei geplanter Wiederverwendung des Tests sogar an Stelle des Ankreuzens im Aufgabenblatt das Ankreuzen auf einem separaten *Anwortblatt* vorsehen.

Zur Aufgabenreihenfolge:

Es empfiehlt sich, die Aufgaben (wenigstens innerhalb eines Blocks) nach steigendem (geschätztem) Anforderungsgrad zu ordnen. Dieses Anordnen vom „Leichten“ zum „Schweren“ vermeidet eine vorzeitige Frustration bei leistungsschwachen Schülern. Außerdem kann der Testbewerter so besser auf Lernlücken schließen. Wurden z. B. alle Wissensfragen gelöst und keine der Verständnisfragen beantwortet, so kann man auf eine unzureichende Verarbeitung des Stoffes schließen. So wird ermöglicht, entsprechende didaktische Maßnahmen zu planen.

Um leistungsstarke Lerner lange genug zu beschäftigen, können an den Schluss des Tests einige zeitaufwendige Zusatzaufgaben gesetzt werden, die außerhalb der Bewertung liegen.

Zur Distraktorenreihenfolge:

Die Distraktoren werden so angeordnet, dass die verlangten Lösungen nicht immer an der gleichen Stelle stehen. Im Gesamttest sollten die Lösungen sich gleich häufig auf jede der möglichen Lösungspositionen verteilen.

Lösungsschablonen (nur bei reinen MC-Tests sinvoll):

Nach der Testzusammenstellung kann man eine vorläufige Lösungsschablone anfertigen. Entweder nimmt man einen Antwortbogen (s. Fig. 16) und schneidet die Buchstaben der richtigen Lösung aus oder man schneidet aus einem Mustertest die „Lösungskästchen“ aus. Legt man eine solche Schablone über einen Antwortbogen bzw. die zugehörige Testseite, so erscheinen in den „Fenstern“ die richtigen Ankreuzungen. Sie können so leichter gezählt und auf dem Auswertungsfeld notiert werden. Die endgültige Schablone wird aus Klarsichtpapier angefertigt, damit sich auch ungültige Doppelankreuzungen erkennen lassen.

Antwortbögen (nur bei reinen MC-Tests möglich):

Soll ein Klassensatz von reinen MC-Tests später wiederverwendet werden, so versieht man die Antwortalternativen statt mit „Kästchen„“ mit *Buchstaben* und entwirft ein Blatt, auf dem die Testteilnehmer bei jedem Item die Buchstaben ankreuzen, die der von ihnen gewählten „Lösung“ entsprechen. Änderungen müssen radiert oder entsprechend gekennzeichnet werden.

Ein solcher Bogen ist in Verbindung mit einer Lösungsschablone sehr „auswerterfreundlich“, stellt jedoch erhöhte Anforderungen an die Sorgfalt der getesteten Schüler!

Beispiel:

ANTWORTBOGEN

Name Vorname Punkte

Klasse Datum Prüfungsfach

Streiche den Buchstaben, der deine Antwort kennzeichnet, dick an!
Wenn du z. B. "C" für richtig hältst, so: Nr. 0 A B ~~C~~ D E

1.	A	B	C	D	E	11.	A	B	C	D	E
2.	A	B	C	D	E.	12.	A	B	C	D	E.
3.	A	B	C	D	E.	13.	A	B	C	D	E.
4.	A	B	C	D	E	14.	A	B	C	D	E
5.	A	B	C	D	E	15.	A	B	C	D	E
6.	A	B	C	D	E	16.	A	B	C	D	E
7.	A	B	C	D	E	17.	A	B	C	D	E
8.	A	B	C	D	E	18.	A	B	C	D	E
9.	A	B	C	D	E	19.	A	B	C	D	E
10.	A	B	C	D	E	20.	A	B	C	D	E

Fig. 16

Testanleitung (Instruktion)
Den Prüflingen sollte genau gesagt werden, was sie zu tun haben. Die dazu nötigen genauen Angaben bilden die *Testanweisung*. Diese Anweisung kann
– schriftlich auf dem Testbogen erfolgen oder
– mündlich vor der Testdurchführung gegeben werden.
Wird ein solcher Test von mehreren Schülergruppen bearbeitet, so hat die Testanweisung einen hohen Stellenwert hinsichtlich der Vergleichbarkeit der Testbedingungen.

Testdurchführung
Für die Durchführung objektivierter Leistungsprüfungen gibt man die Testaufgabenblätter (ggf. mit Antwortbögen wie in Fig. 16 mit einer allgemeinen Testanleitung (Arbeitseinweisung zur Handhabung der Testunterlagen) an alle Teilnehmer aus und startet nach dem Eintragen der Schülerdaten die Bearbeitungsphase.

Am Ende der Bearbeitungsphase werden alle Unterlagen (insbesondere die Antwortbögen, falls solche ausgegeben wurden) wieder gesammelt.

Analyse von lehrzielorientierten informellen Tests.

Aufgabenanalyse:

Eine empirische Aufgabenanalyse soll objektive Auskünfte zu *Schwierigkeitsgrad* und *Trennschärfe* der Items, gegebenenfalls auch die Eignung der Distraktoren bei MC-Items liefern. Testtheoretisch dienen die so gewonnenen Daten der *Bewertung* (Evaluation) und Verbesserung von Tests. Denn ungeeignete Aufgaben können erkannt, später verbessert oder sofort zur Ermittlung der endgültigen Punktzahl ausgeschieden werden. Aufgabenanalysen geben darüber hinaus wichtige Auskünfte zur Untersuchung von Lehr-Lern-Prozessen.

Man sollte dabei nicht nur von den ermittelten „Lösungsquoten“ der Testaufgaben nach einem Unterricht auf ihre Eignung schließen. Aufgaben, die nur wenig gelöst wurden, können ja auch Lehrschwächen andeuten. Sie können Unzulänglichkeiten der vorausgegangenen Lehrsequenz aufdecken.
Den angesprochenen Zielsetzungen soll nun nachgegangen werden.

Vorbereitung der Analyse:
Nach Auswertung der Antwortbögen bzw. Testblätter (d.h. der Feststellung einer vorläufigen Punktzahl) kann die Aufgabenanalyse in Angriff genommen werden. Dazu ordnet man alle Bearbeitungen nach steigender Punktzahl und teilt sie eine leistungsschwächere Hälfte (Untergruppe U) und eine leistungsstärkere Hälfte (Obergruppe O) auf.

Bei ungerader Anzahl von Teilnehmern bleibt ein Antwortbogen der Mittelgruppe bei der Analyse unberücksichtigt.

Danach erstellt man ambesten ein *Analyseblatt*. Auf diesem Blatt sollen alle Daten gesammelt werden, die für eine Analyse notwendig sind. Es enthält die *Nummern aller Testaufgaben*, bei MC-Items die *Antwortangebots*bezeichnungen und leere Spalten mit den folgenden Benennungen:

Aufgabenanalyse

Klasse:____________ Datum: ____________ Fach:____________

Aufgabe	Antwortangebote (Lösung eingekreist)					Lösungsanzahl in den Gruppen				Aufgabenbewertung			
	A	B	C	D	E	O	U	O+U	O–U	zu leicht	zu schwer	nicht trennscharf	nicht gewählte Distaktoren
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													

Fig. 17

Darin bedeuten:

O = Anzahl der richtigen Lösungen in der Obergruppe („bessere“ Hälfte)
U = Anzahl der richtigen Lösungen in der Untergruppe („schwächere“ Hälfte)
O+U = Gesamtzahl der richtigen Lösungen
O+U = Differenz der Anzahl richtger Lösungen zwischen Obergruppe und Untergruppe

Durchführung der Analyse bei MC-Items:
Jedes Item i wird unter den drei Gesichtspunkten

1. Wie hoch ist sein *Schwierigkeitsgrad* q_i ?
2. Wie hoch ist sein *Trennschärfeindex* t_i ?
3. Sind seine Distraktoren geeignet?

untersucht. Bei anderen Aufgabenformaten entfällt der dritte Punkt. Dafür ist vor der Definition des Schwierigkeitsgrads und des Trennschärfeindex eventuell eine „Umkodierung der Bewertung“ nötig. So ist bei einem Item mit 4 möglichen Wertungspunkten zu überlegen, ob man es ab 3 Punkten oder erst ab 4 Punkten bei der Analyse als „richtig“ gelöst ansehen will!

Wir übernehmen Vorschläge aus Beiner (1982), die für Schulpraktiker handhabbar sind:

Zu 1.: Unter dem *Schwierigkeitsgrad* q_i eines Items/einer Aufgabe i soll die relative Quote der getesten Schüler verstanden werden, die an der Aufgabe *gescheitert* sind. Haben z.B. 35 % der Schüler die Aufgabe i nicht richtig gelöst, so ist $q_i = 0,35$ $(= 35\ \%)$.

Ein hoher Schwierigkeitsgrad einer einzelnen Testaufgabe liegt ja vor, wenn sie nur von wenigen Testpersonen richtig gelöst wurde[4]

In plakativer Form sich die Bestimmung so:

$$\text{Schwierigkeitsgrad } q = \frac{\text{Anzahl der richtigen Antworten}}{\text{Anzahl der Prüflinge}} = 1 - \frac{O+U}{N}$$

Auf diese Weise entspricht ein Wert nahe bei 1, also ein relativ hoher Wert, auch tatsächlich einem hohen Schwierigkeitsgrad, umgekehrt ein niedriger Wert auch einer leichten Aufgabe. Die Bestimmung der Schwierigkeitsgrade dient

a) der Kontrolle, ob die Aufgaben und der Test für die Adressaten angemessen war (mittlere Schwierigkeiten gelten oft als pädagogisch und messtheoretisch vorteilhaft)

b) der Identifizierung möglicher Konstruktionsfehler in Aufgaben.

Aufgaben, die von fast allen Kandidaten richtig gelöst wurden, tragen nämlich zur Unterscheidung der leistungsstärkeren von den leistungsschwächeren Personen genauso wenig beitragen, wie Aufgaben, die von fast niemandem richtig gelöst wurden.
Eine gängige Empfehlung lautet, dass geeignete Testaufgaben q-Werte von 0,15 und 0,85 haben sollten. Wer nicht mit Brüchen und Relativquoten rechnen will, ermittelt für eine gesamte Testung die Grenzwerte $S = 0,15 \cdot N$ und $L = 0,85 \cdot N$ der Lösungsanzahlen. Dann gilt ein Item

als zu schwer, wenn $O+U < S$,

als zu leicht, wenn $O+U > L$

Beispiel:
40 Schüler haben eine Aufgabe bearbeitet. Dann sollte die Anzahl $O+U$ der richtigen Lösungen (einschließlich der Grenzen) zwischen $0,15 \cdot 40 = 6$ *und* $0,85 \cdot 40 = 34$ *liegen. Haben weniger als 6 Schüler die Aufgabe richtig, so gilt sie als zu schwer. Haben mehr als 34 Schüler die Aufgabe richtig, dann gilt sie als zu leicht.*

Zu 2.: Die *Trennschärfe* gibt an, wie gut eine Aufgabe geeignet ist, zwischen den Leistungsstarken und den Leistungsschwachen (Ober- und Untergruppe) zu unterscheiden. Trennscharfe Aufgaben sind also Aufgaben, deren Ergebnis in hohem Maße mit dem gesamten Testergebnis übereinstimmt. Das heißt, wer eine trennscharfe Aufgabe richtig löst, ist mit hoher Wahrscheinlichkeit im ganzen Test erfolgreich.

[4] In der üblichen Literatur zu informellen Tests findet sich leider die *Lösungsquote* als Bestimmungsgröße, die eher „Leichtigkeitsgrad“ genannt werden sollte. Dort wird nämlich der „Schwierigkeitsindex“ einer Aufgabe als Anteil p_i der richtigen Antworten an der der Gesamtzahl N der Antworten bestimmt. Der Schwierigkeitsindex kann somit Werte zwischen 0 und 1 annehmen. Schwierigen Aufgaben entspricht bei dieser Berechnung aber ein niedriger Zahlenwert und leichten Aufgaben ein hoher Wert nahe 1. Als Formel wird dabei verwendet $p = \frac{O+U}{N}$ mit N =Anzahl der der Prüflinge.

Der *Trennschärfe-Index* t_i errechnet sich mit den Werten aus dem Analysebogen nach der Formel:

$$t = \frac{O - U}{\frac{1}{2}N}$$

Die Trennschärfe sollte möglichst hoch sein, aber auf jeden Fall positiv. Ein negativer Trennschärfe-Index würde sich ergeben, wenn eine Aufgabe überwiegend von Teilnehmern aus der Untergruppe gelöst würde, nicht aber von Schülern der Obergruppe. Die Differenz $O - U$ wäre dann negativ und die Aufgabe somit unbrauchbar, da ein negativer Zusammenhang von Einzelaufgabenergebnis und Gesamttestergebnis bestehen würde. (So etwas deutet auf Unklarheiten und/oder auf fehlende Validität hin.)

Beispiele:

t	=	$0,80$	*Die Aufgabe wurde überwiegend von der leistungsstarken Gruppe gelöst; denn $O - U$ ist deutlich größer als Null.*
t	=	0	*Die Aufgabe wurde von beiden Gruppen gleich oft gelöst; denn $O - U$ ist gleich Null.*
t	=	$-0,20$	*Die Aufgabe wurde öfter von der Untergruppe gelöst; denn $O - U$ ist kleiner als Null.*

Aufgaben sollen als ungeeignet angesehen werden, wenn der Trennschärfe-Index kleiner als 0,1 ist (also insbesondere auch, wenn er negativ ist).

zu 3.: Als *Distraktoren* bezeichnet man (wie bereits erwähnt) die *Falschantworten* bei MC-Items. Diese Antworten müssen so beschaffen sein, dass sie zufällige Lösungen der Aufgabe möglichst verhindern. Gute Distraktoren sind solche, die häufig und fast gleich oft wie die richtigen Lösungen gewählt werden. Sie sollten aus dem Bereich der Falschantworten kommen, die man bei offener Aufgabenform bekommen würde. Unter diesen Bedingungen am häufigsten vorgetragene Falschantworten dürften geeignete Distraktoren sein.

Die Eignungsfeststellung von Distraktoren geschieht am besten mit Hilfe des Blattes „Aufgabenanalyse“. Distraktoren, die wenig oder gar nicht gewählt werden, haben keine Anziehungskraft. Um sie nicht zu wählen, braucht man keinen Sachverstand.

Vergleichen wir die Aufgabenanalye an einem Beispiel von 10 Aufgaben mit 22 Testteilnehmern. Welche Aufgaben des Tests sind noch nicht voll brauchbar?

Als Grenzwerte würde man hier mit $N = 22$ wählen:

$0,85N$	=	$18,7;$	d.h.:	zu leicht, wenn $O + U \geq 19$
$0,15N$	=	$3,3;$	d.h.:	zu schwer, wenn $O + U \leq 3$
$0,05N$	=	$1,1;$	d.h.:	nicht trennscharf, wenn $O - U \leq 1$

Demnach wären im folgenden Analysebogen (Fig. 18) die Aufgaben 3 und 9 nicht trennscharf, die Aufgabe 5 zu leicht und die Aufgabe 10 zu schwer.

Dass bei der zu leichten Aufgabe ein Distraktor nicht gewählt wurde, spricht insofern noch nicht gegen die Qualität der Distraktoren, als es nur drei Nichtlöser der Aufgabe und vier Distraktoren gibt.

Verdächtiger ist der nicht gewählte Distraktor der dritten Aufgabe, da diese bei „mittlerer Schwierigkeit“ eine unzureichende Trennschärfe besitzt.

Man würde bei diesem Test die Interpretation der Resultate möglichst nur auf die Aufgaben 1, 2, 4, 6, 7 und 8 stützen, wenn diese noch die intendierten Lehrziele vertreten.

Aufgabenanalyse

Klasse: 8 a Datum: 19.12.05 Fach: Mathematik

Aufgabe	Antwortangebote (Lösung eingekreist) A	B	C	D	E	Lösungsanzahl in den Gruppen O	U	O+U	O–U	Aufgabenbewertung zu leicht	zu schwer	nicht trennscharf	nicht gewählte Distraktoren
1	4	10	3	2	3	7	3	10	4				
2	3	5	2	9	3	8	1	9	7				
3	12		2	4	3	6	6	12	0			X	B
4	1	5	1	2	13	9	4	13	5				
5	1	1	19		1	11	8	19	3	X			D
6	3	2	13	2	2	8	5	13	3				
7	1	17	2	1	1	12	5	17	7				
8	3	4	3	8	4	6	2	8	4				
9	1	6	3	1	11	6	5	11	1			X	
10	2	5	4	5	6	2	0	2	2		X		

Fig. 18

Prüfung der Testgüte

Der Forderung nach hoher Test-Güte entsprechend wurden schon einige Konstruktionsschritte empfohlen, die wir nun für einen LIT als realisiert ansehen. Solche Maßnahmen können nach der Testvorlage ergänzt werden durch Verbesserung und Ergänzung der Aufgaben auf der Basis von Schülerbefragungen, Expertenurteilen und neuen Formulierungen.

Wie eine zusätzliche statistische Überprüfung der Gütekriterierin aussehen kann, soll im Folgenden anhand von Beispielen vorgeführt werden. Wir benutzen dafür den *Übereinstimmungskoeffizienten* von FRICKE, mit dem z.B. die Übereinstimmung zwischen mehreren dichotomen Expertenurteilen oder Prüferurteilen überprüft werden kann[5]:

$$\text{Ü} = \frac{\text{beobachtete Übereinstimmung}}{\text{maximal mögliche Übereinstimmung}} = 1 - \frac{4 \cdot (k \cdot \sum x_i - \sum x_i^2)}{N \cdot k^2}$$

mit

[5] d.h. jedes Urteil lautet entweder 0 oder 1

k = Anzahl der Datenreihen, die auf Übereinstimmung geprüft werden sollen
N = Anzahl der Probanden
x_i = Summe der Messwerte, die zu dem Probanden i gehören
$\sum$ bedeutet die Summenbildung über alle Probanden

Dieser Koeffizient ist so angelegt, daß er den Zahlenwert 1 einnimmt, wenn maximale Übereinstimmung festzustellen ist, und den Wert 0 bei minimaler Übereinstimmung.

Beispiel zur *Objektivitätsprüfung*:
4 Prüfer haben unabhängig voneinander anhand der vorgelegten Testergebnisse entschieden, wer von 10 Prüflingen das getestete Lehrziel erreicht hat (Urteil=1) oder verfehlt hat (Urteil=0):

		Schüler $(N=10)$										
		1	2	3	4	5	6	7	8	9	10	
Beurteiler $(k=4)$	(1)	0	1	1	1	1	0	1	1	0	0	
	(2)	0	1	1	0	1	0	0	1	0	0	
	(3)	0	1	1	1	1	0	1	1	0	1	
	(4)	0	1	1	1	1	0	1	1	0	0	
Pb-Summe	x	0	4	4	3	4	0	3	4	0	1	$\sum x = 23$
	x^2	0	16	16	9	16	0	9	16	0	1	$\sum x^2 = 83$

$$\ddot{U} = 1 - \frac{4 \cdot (4 \cdot 23 - 83)}{10 \cdot 16} = 0,78$$

Bei der Betrachtung der Tabelle fällt auf, dass die Beurteiler bei sieben Prüflingen (1; 2; 3; 5; 6; 8; 9) völlig übereinstimmen.

Bei drei Prüflingen (4; 7; 10) gibt es mäßige bis gute Übereinstimmung (3 von 4 Prüfern stimmen noch überein).

Die Objektivität dieser Schülerbeurteilung ist demnach mäßig bis gut wie auch der Übereinstimmungskoeffizient mit dem Wert $0,78$ zeigt.

Beispiel zur Zuverlässigkeitsbestimmung nach einem zweifachen Testeinsatz (Test und Re-Test), bei dem jedesmal nach einem objektiven Schlüssel das Urteil „Lehrziel erreicht“ (1) oder „Lehrziel nicht erreicht“ (0) zu vergeben war:

		Schüler $(N=10)$										
		1	2	3	4	5	6	7	8	9	10	
Test	(1)	1	0	1	1	0	1	0	1	1	1	
Retest	(2)	1	1	1	1	0	0	0	1	0	1	
Pb-Summe	x	2	1	2	2	0	1	0	2	1	2	$\sum x = 13$
	x^2	4	1	4	4	0	1	0	4	1	4	$\sum x^2 = 23$

$$\text{Ü} = 1 - \frac{4 \cdot (2 \cdot 13 - 23)}{10 \cdot 4} = 0,70$$

Die Zuverlässigkeit als Wiederholbarkeit ist bei diesem Test mäßig; nur bei fünf Schülern stimmen die Testdaten mit den Retestdaten überein.

Wenn bei einem Test-Retest-Paar Testpunktzahlen zu vergleichen sind, verwendet man an Stelle des Übereinstimmungskoeffizienten den *Korrelationskoeffizienten*:

$$\rho_{x,y} = \frac{N\sum x \cdot y - (\sum x)(\sum y)}{\sqrt{(N\sum x^2 - (\sum x)^2) \cdot (N\sum y^2 - (\sum y)^2)}}$$

Dieser Wert wäre in unserem letzten Beispiel nur etwa 0,36 !

Endgültige Testwertermittlung und Interpretation

Nach einer Aufgabenanalyse steckt die Testgüteprüfung einen gewissen Rahmen für die Gewinnung endgültiger Testwerte und deren Deutung (Interpretation) ab. In diesem Zusammenhang sei vor zu starker „Testgläubigkeit“ gewarnt. Es sollte klar sein, dass es immer Messfehler gibt. Deshalb sollten die Testergebnisse möglichst nicht einzeln, sondern im Zusammenhang mit mehreren Tests und Prüfverfahren im Sinne von *Trendaussagen* gedeutet werden.
Zur Ermittlung der endgültigen Punktzahlen werden die folgenden ungeeigneten Aufgaben ausgeschlossen (falls noch genug Aufgaben für das Begründen von Urteilen übrig bleiben!!!):

1. zu leichte Aufgaben (q kleiner als 15 %, d.h. p größer als 85 %)

2. zu schwere Aufgaben (q größer als 85 %, d.h. p kleiner als 15 %)

3. Aufgaben mit geringer Trennschärfe (t kleiner als 10 %, d.h. $O-U$ kleiner als 5 % von N)

4. Aufgaben mit ungeeigneten Distraktoren (zu selten gewählt).

Anhand einer neuen Auszählung, die nur die geeigneten Aufgaben berücksichtigt, kann nun die endgültige Punktzahl ermittelt werden. Sie bildet die Grundlage für die nachfolgende Interpretation des Testergebnisses:.
Dient der Test hauptsächlich der *Leistungsdifferenzierung* (Einstufungs/Selektionsfunktion), so wird das Testergebnis (endgültige Punktzahl) anhand eines vor der Testdurchführung festgesetzten Schlüssels benotet oder interpretiert.
Hat der Test mehr pädagogische Funktionen, wie Aufdecken von Leistungsschwächen und Fehlern im Lehrprozess, so müssen die Daten zum Schwierigkeitsgrad der Aufgaben genauer untersucht werden: Haben sich z.B. sehr viele Aufgaben auf den höheren Anforderungsebenen als zu schwierig erwiesen, so muss das nicht unbedingt mangelnde Eignung dieser Aufgabe bedeuten. Es kann auch darauf hindeuten, dass der Lehrer das Unterrichtsziel – bis zur Ebene des Problemlösens vorzudringen – nicht erreicht hat. Solche *Unterrichtslücken* müssten dann im nachfolgenden Unterricht aufgearbeitet werden. Eventuell müsste der Lehrer auch seine Lehrmethoden überprüfen.

II.4 Leistungsmessung und Notengebung

Wie lässt sich eine Lösungswahrscheinlichkeit p für glrichartige Aufgaben (bzw. eine mittlere Lösungswahrscheinlichkeit für eine *Aufgabenpool* in *Noten* umwandeln?
Das Problem besteht offensichtlich in der dazu notwendigen Unterteilung des Intervalls $]0;1[$.
Man kann trefflich darüber streiten, ob die Teilpunkte

$$0,4,\ 0,55,\ 0,7,\ 0,85$$

der einheitlichen Bewertungsempfehlungen der KMK oder eine Unterteilung des Typs

$$0,5,\ 0,75,\ 0,875,\ 0,9375$$

„angemessen“ sind. Dies liegt daran, dass solche Vorschläge von der Art der geprüften Fertigkeit abhängen.
Die KMK-Empfehlung gilt für schriftliche Prüfungsleistungen, die in Form komplexer Aufgaben mit mehrstufiger Bewertung erbracht wurden. Würde man eine solche Aufgabe nur mit 0 oder 1 bewerten, so ist wohl eine „Lösungswahrscheinlichkeit“ von 0,4 in den meisten Fällen tatsächlich „gerade noch ausreichend“.
Für Routinefertigkeiten scheint der zweite Vorschlag eher passend, dürfte jedoch z. B. für das Beherrschen des kleinen 1×1 noch nicht extrem genug sein.
Um eine einheitliche Behandlung solcher Fragen zu ermöglichen, betrachten wir erst einmal das Problem, wie man Lösungswahrscheinlichkeiten „multiplikativ“ vergleicht.

Physikalisch ist *Leistung* der Quotient *Arbeit*/*Zeit*, in den Augen von Mathematiklehrer(innen) aber eher das Erreichen einer *Niveaustufe*, die sich eigentlich nicht durch die Anhäufung geleisteter gleichartiger „Arbeitsschritte“ definieren lässt. Wir schließen uns dieser Sichtweise zunächst nicht an und vergleichen in einem Besipiel zwei Schüler, die unterschiedliche „Arbeitsgeschwindigkeiten“ haben.

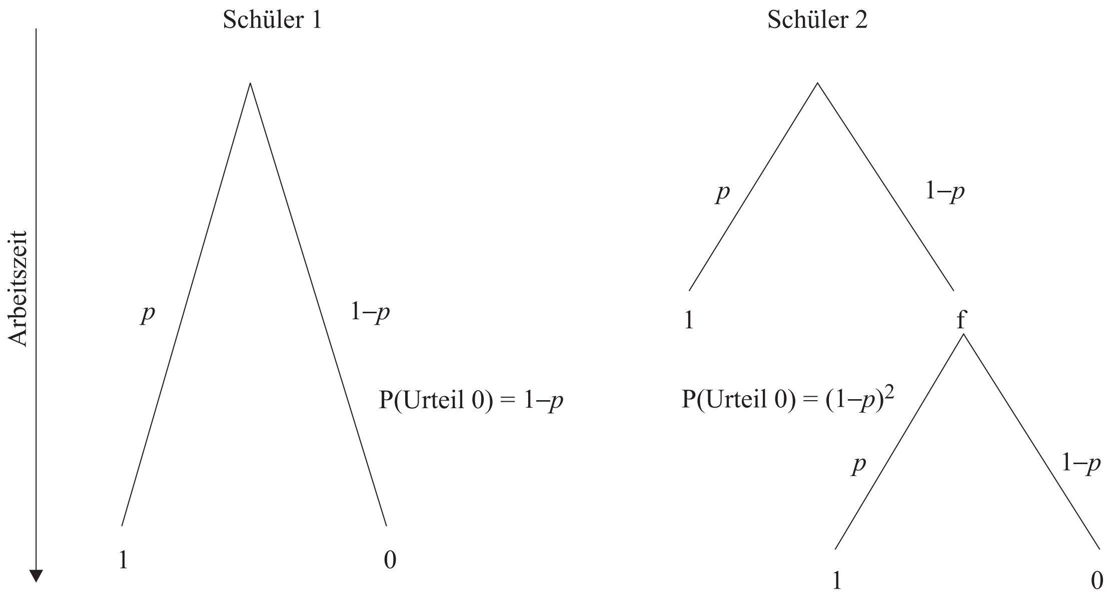


Fig. 19

Unter der Annahme, dass Schüler 2 doppelt so schnell wie Schüler 1 ist, die gleiche Lösungswahrscheinlichkeit p bei einem Einzelversuch wie Schüler 1 hat, und den freien Zeitblock für

das Erkennen und Reparieren eines eventuellen Fehlers im ersten Zeitblock nutzt, beträgt die Wahrscheinlichkeit für das endgültige Versagen von Schüler 2 nur $(1-p)^2$, während Schüler 1 mit der Wahrscheinlichkeit $1-p$ versagt. Dies würde es nahe legen, Schüher 2 „doppelt so sicher“ wie Schüler 1 zu nennen.
Wir verallgemeinern das Beispiel und ersinnen ein Prüfverfahren, das vom Vergleich der Arbeitszeiten unabhängig macht:

Redundanzmodell:

Angenommen, ein Schüler k wird folgendermaßen geprüft:

(1) Der Schüler k erhält eine zufällig gewählte Aufgabe a_1. Löst er sie, so wird als Versuchsergebnis eine 1 notiert und die Prüfung ist beendet.

(2) Löst k die Aufgabe a_i (beim ersten Mal ist i gleich 1) nicht, so wird eine Aufgabe a_{i+1} zufällig gewählt. Löst k diese Aufgabe, so wird die Nummer $i+1$ als Versuchsergebnis notiert und die Prüfung ist beendet. Ansonsten wird wieder eine Aufgabe präsentiert.

Je mehr Versuche ein Schüler braucht, desto leistungsschwächer erscheint er in der Prüfung.

Hat nun ein Schüler k bei der ersten Aufgabe dieselbe Chance für den Prüfungserfolg, wie ein Schüler j bei bei insgesamt n Aufgaben, so kann man k „*n*-mal so sicher“ wie j nennen.

Definition II.1

Ein Schüler *b* heißt bezüglich einer Aufgabe **n-mal so sicher** wie ein Schüler a

$$:\Longleftrightarrow$$

Für die Lösungswahrscheinlichkeiten $p^{(a)}$ und $p^{(b)}$ dieser Schüler für diese Aufgabe gilt

$$1-p^{(b)} \;=\; (1-p^{(a)})^n \,.$$

Folgerung: Man erhält n in der Definition aus der Beziehung $1-p^{(b)} = (1-p^{(a)})^n$ durch Logarithmieren in der Form

$$n = \frac{\ln(1-p^{(b)})}{\ln(1-p^{(a)})} \,.$$

Setzt man die so definierte Funktion auf $]0;1[^{\times 2}$ fort, so erhält man ein Vergleichsmaß für Lösungswahrscheinlichkeiten:

Definition II.2

Haben Schüler a und b für eine Aufgabe i die Lösungswahrscheinlichkeiten p_a bzw. p_b mit $p_a, p_b \in]0;1[$, so heißt

$$\rho_i(a,b) := \frac{\ln(1-p_b)}{\ln(1-p_a)}$$

das **Sicherheitsverhältnis** von b zu a bezüglich Aufgabe i.

Offensichtlich gilt $\rho_i(c,b)\cdot\rho_i(b,a) = \rho_i(c,a)$ für alle Schüler a,b,c einer Population.
Logarithmiert man das Sicherheitsverhältnis noch einmal, so erhält man ein additives Maß:

Definition II.3

Haben Schüler a und b für eine Aufgabe i die Lösungswahrscheinlichkeiten p_a bzw. p_b mit $p_a, p_b \in]0;1[$, so heißt

$$d_i(a,b) := \ln(\frac{\ln(1-p_b)}{\ln(1-p_a)}) = \ln(-\ln(1-p_b)) - \ln(-\ln(1-p_a))$$

die **Distanz** von b zu a bezüglich Aufgabe i.

Während das Sicherheitsverhältnis von der Wahl der Logarithmenbasis *unabhängig* ist, würde sich in Definition 2.II.3 ein anderes Maß ergeben, wenn an Stelle des natürlichen Logarithmus z.b. der dekadische Logarithmus verwendet worden wäre.
In dem Beispiel mit der Unterteilung des Intervalls $]0;1[$ durch die Teilpunkte $0,5; 0,75; 0,875; 0,9375,\ldots$ ergibt sich für Schüler a, b und c mit $p_a = 0,5$, $p_b = 0,75$ und $p_c = 0,9375$ (bezüglich einer Aufgabe) aus der letzten Definition

$$d(b,a) = \ln 2 \approx 0,693,\ d(c,b) = \ln 2 \approx 0,693,\ d(c,a) = \ln 3 \approx 1,386\ .$$

Nach Wahl einer Lösungswahrscheinlichkeit p_0, die den Nullpunkt auf der Distanzskala festlegt, ergibt sich eine Transformation η des Intervalls $]0;1[$ nach $\mathbb{R}$, deren Vorschrift z.b. für $p_0 = \frac{1}{2}$ durch

$$\eta(p) := \ln(-\ln(1-p))\ -\ \ln(\ln(2))$$

gegeben ist. Wie der Graph von η zeigt, ist diese Transformation im Teilintervall $]0,3;0,7[$ noch näherungsweise linear. In der Nähe der Intervallenden wird dagegen die *Spreizung* des Intervalls $]0;1[$ sehr deutlich:

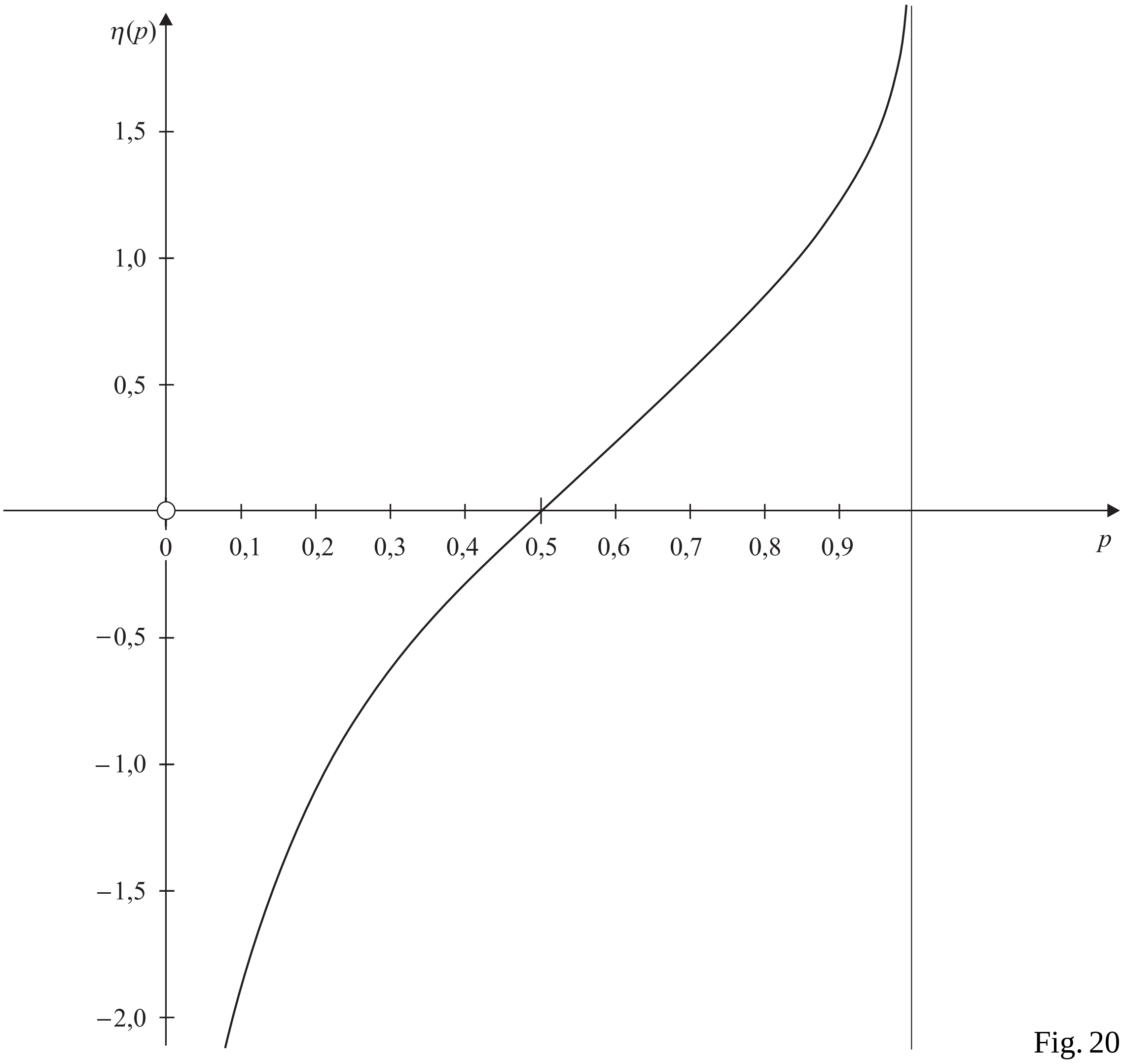


Fig. 20

Der Graf ist zwar „S-förmig“, jedoch nicht punktsymmetrisch bezüglich des Punktes (0,5;0). Er erinnert an die Transformation mit der Vorschrift

$$\kappa : \; p \mapsto \frac{2}{\pi} \arcsin\left(\sqrt{p}\right),$$

die von K. J. KLAUER 1982[6] zur Definition von Teilpunkten des Einheitsintervalls mit der Begründung vorgeschlagen worden war, dass damit in allen Leistungsbereichen in etwa die Schätzunsicherheit des transformierten Wertes gleich groß wird.

Da die Schätzunsicherheit wohl kein Maßstab für theoretische Leistungsunterschiede sein kann, diskutieren wir den KLAUERschen Vorschlag erst einmal nicht und wenden uns der Frage zu, wie man Distanzen linear in die üblichen Noten umrechnen kann. Offensichtlich reicht es dabei, (mittlere) Lösungswahrscheinlichkeiten p_u und p_o mit $p_u < p_o$ festzulegen, bei denen (vor der Rundung auf die üblichen Notenstufen) die Noten 4,49 bzw. 1,49 zu geben sind. Daraus resultiert dann eine Notenfunktion mit der Vorschrift

[6] K.J. Klauer, Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 1982, *14*, S. 65-79

$$(2.4.1) \qquad \nu(p) := b - a \cdot \ln(-\ln(1-p)) \qquad (0 < p < 1)$$
$$\text{mit } b = 4,49 + \frac{3\ln(-\ln(1-p_u))}{\ln(-\ln(1-p_o)) - \ln(-\ln(1-p_u))}$$
$$\text{und } a = \frac{3}{\ln(-\ln(1-p_o)) - \ln(-\ln(1-p_u))} \, .$$

Beispiel (Routinefertigkeit):

Angenommen, für eine Routinefertigkeit wird ein „mittlerer" Beherrschungsgrad (das heißt $p_u = 0,5$) als *noch ausreichend* angesehen. Einem Lerner u mit der Lösungswahrscheinlichkeit p_u soll also die Note 4,49 gegeben werden. Soll einem Lerner o mit der Lösungswahrscheinlichkeit $p = p_o := 0,9375$ noch die Note *sehr gut* gegeben werden, so muss die Notenfunktion mit der Vorschrift

$$\nu(p) := 3,697 - 2,164\ln(\ln(1-p))$$

verwendet werden. Würde man an Stelle des natürlichen Logarithmus mit dem dekadischen Logarithmus log arbeiten, so hätte die Vorschrift die Form

$$\nu(p) := 1,892 - 4,893\log(\log(1-p)) \, .$$

Im Falle $\nu(p) > 6$ ist die Note 6 zu erteilen. Falls sich $\nu(p) < 1$ ergibt, ist die Note 1 zu erteilen. Für die Notenstufen ergeben sich die (gerundeten!) Grenzen in folgender Tabelle:

Note	5,49	4,49	3,49	2,49	1,49
$p_{\min}$	0,35	0,50	0,67	0,83	0,94

Beispiel (KMK-Stufen):

Setzt man $p_u = 0,5$ und $p_o = 0,85$, so ergibt sich die Notenfunktion mit der Vorschrift

$$\nu(p) := 2,954 - 2,286\ln(\ln(1-p)) \, .$$

Für die Notenstufen ergeben sich gerundet nahezu die Grenzen des KMK-Schlüssels:

Note	(5,49)	4,49	3,49	2,49	1,49
$p_{\min}$	(0,28)	0,50	0,55	0,71	0,85

Die Grenze zwischen den Noten *mangelhaft* und *ungenügend* ist eingeklammert, da sie in den ursprünglichen Richtlinien nicht angesprochen war.

Würde man die Transformation κ von KLAUER heranziehen, so hätte man bis auf geringe Abweichungen fast die gleichen Notenschlüssel wie bei der Verwendung von η.