

bifie | Bildung

 standards

Item Entwicklung: Kognitive Prozesse und Item Schwierigkeit

*Roman Freunberger
Ursula Itzlinger-Bruneforth*



Bundesinstitut für Bildungsforschung, Innovation & Entwicklung
des österreichischen Schulwesens
Alpenstraße 121 / 5020 Salzburg

www.bifie.at

Item Entwicklung: Kognitive Prozesse und Item Schwierigkeit

BIFIE | Department Bildungsstandards & Internationale Assessments (BISTA),
Salzburg 2013

Der Text sowie die Aufgabenbeispiele können für Zwecke des Unterrichts in österreichischen Schulen sowie von den Pädagogischen Hochschulen und Universitäten im Bereich der Lehreraus-, Lehrerfort- und Lehrerweiterbildung in dem für die jeweilige Lehrveranstaltung erforderlichen Umfang von der Homepage (www.bifie.at) heruntergeladen, kopiert und verbreitet werden. Ebenso ist die Vervielfältigung der Texte und Aufgabenbeispiele auf einem anderen Träger als Papier (z. B. im Rahmen von Power-Point Präsentationen) für Zwecke des Unterrichts gestattet.

Autorinnen und Autoren:

Roman Freunberger
Urusla Itzlinger-Bruneforth

Inhaltsverzeichnis

3 1 Einleitung

4 2 Die Erhebungs-Triade

5 3 Kognitive Modelle und Itementwicklung

5 3.1 Der Konstruktrahmen

6 3.2 Itemmodelle

7 3.3 Schwierigkeitsbestimmende Merkmale von Items

11 4 Exkurs

11 4.1 Lernzieltaxonomien

11 4.1.1 Blooms Taxonomie

12 4.1.2 Überarbeitung der Bloom'schen Taxonomie

13 4.1.3 Eine neue Taxonomie (NT)

15 4.2 Gedächtnissysteme

17 Literatur



1 Einleitung

Die Entwicklung eines Testinstruments zur Kompetenzdiagnostik beinhaltet eine enge Verknüpfung von statistischen Modellen und kognitiver Psychologie (Gorin, 2006; Mislevy, 2006; Rupp & Templin, 2008). Eine wichtige Fragestellung steht hier im Zentrum, und zwar welche kognitiven Operationen an der Lösung eines Items beteiligt sind (Dimitrov & Raykov, 2003). Im Grunde geht es darum festzustellen, welche kognitiven Verarbeitungsprozesse die Schwierigkeit von Items am besten vorhersagen können¹. Laut Dimitrov und Raykov (2003) ist dies hilfreich, um Items mit einem bestimmten Schwierigkeitsgrad zu entwickeln, ohne diese Items vorher getestet zu haben. So sollen Kompetenzstufenmodelle und auch Lehrstrategien entwickelt werden, die gezielt bestimmte Kompetenzen mit unterschiedlichem Komplexitätsgrad fördern (Rupp, Leucht & Hartung, 2006).

Nachfolgend werden die Vorarbeiten kurz zusammengefasst, die für den Itemwriting-Prozess relevant sind. Vor allem Itemmodelle, welche schwierigkeitsbestimmende Merkmale erfassen, sind hier von besonderer Wichtigkeit. Eine allgemeine Anleitung für die Entwicklung von Items mit variierendem Komplexitätsgrad kann hier nicht gegeben werden. Dies ist, und das soll in diesem Rahmen ausgeführt werden, immer abhängig von dem jeweiligen Konstrukt, das überprüft werden soll. Der Schwierigkeitsgrad von Items muss daher immer im Kontext der jeweiligen Leistungsmessung, also der Fragestellung und der zu untersuchenden Population, gesehen werden. Möchte man zum Beispiel die Lesefähigkeit bei Schülerinnen und Schülern auf der vierten Schulstufe überprüfen, muss zuerst in einer Konstruktbeschreibung (z. B. den Bildungsstandards) definiert werden welche Kompetenzen hier gefragt sind. Des Weiteren kann festgelegt werden, welche Items dieses Konstrukt messen können, wie diese gestaltet werden sollen und welche Aufgabenmerkmale den Grad der Komplexität spezifisch für das zu messende Konstrukt beeinflussen können. In anderen Worten, es sollten Items generiert werden, die es ermöglichen, zwischen Schülerinnen und Schülern mit hoher und jenen mit geringer Kompetenz in einem bestimmten Kompetenzbereich zu differenzieren.

¹ Methodisch kann dies z. B. durch eine multiple Regression, das linear-logistische Testmodell (LLTM) oder mittels Strukturgleichungsmodellen erfasst werden.

2 Die Erhebungs-Triade

Eine Schulleistungsuntersuchung (engl.: Large-scale-assessment) steht im Wesentlichen auf drei Säulen (vgl. Pellegrino & Chudowsky, 2003), (1) Kognition, (2) Beobachtung und (3) Interpretation. Diese drei Säulen stehen in Verbindung und sind für die Entwicklung von Tests, aber auch von Lehrstrategien von wesentlicher Bedeutung.

Die erste Säule der Kognition umfasst theoretische Modelle, welche die Frage behandeln, wie Personen Kompetenzen in einer bestimmten Domäne erwerben, wie sie Wissen repräsentieren und wie sie dieses Wissen anwenden können. Kognitive Modelle spielen vor allem für die Konstruktbeschreibung, für die Itementwicklung und für die Validierung von Testinstrumenten eine große Rolle (Gorin, 2006).

Im Weiteren stellt die Säule der Beobachtung die eigentliche Erfassung bzw. Erhebung dieser erlernten, kognitiven Kompetenzen dar. Kompetenzen sind im Grunde latente Größen, die nicht direkt beobachtbar sind und daher über sogenannte Indikatoren, die Testitems, gemessen werden (Hartig & Klieme, 2006). Diese Indikatoren oder Items sollten zumindest theoretisch jeweils eine spezifische Kompetenz erfassen. Die Beobachtung von Verhalten erfolgt demnach aufgrund von entwickelten Items, welche bestmöglich die vorhandenen Kompetenzen von Schüler/innen hervorrufen können. Welche Kompetenzen überprüft werden, ist im Konstrukt festgehalten, dieses kann eine kurze Beschreibung oder auch eine Auflistung sein, welche die Kompetenzen in einer bestimmten Domäne beinhaltet.

Gorin (2006) weist darauf hin, dass Konstrukte oft wenig detailliert und sehr allgemein beschrieben sind und daher für die Itementwicklung häufig eine Herausforderung darstellen. Das drückt wiederum die Wichtigkeit von kognitiven Modellen aus, welche meist detailliert beschreiben, was für ein Wissen und welche Fertigkeiten und Fähigkeiten in einer bestimmten Domäne verlangt werden.

Unter dem Aspekt der Interpretation versteht sich schließlich die Analyse (meist durch statistische Testmodelle) dieser erhobenen Kompetenzen und deren formale Beschreibung.

Für die Entwicklung von Tests und Testitems sind alle Säulen wesentlich, es soll hier allerdings verstärkt darauf eingegangen werden, was Items ausmacht, die für die Schüler/innen ansprechend sind und ihre Kompetenzen und ihr Leistungsniveau bestmöglich erfassen können.

3 Kognitive Modelle und Itementwicklung

Für die Itementwicklung und generell für die Entwicklung von Testinstrumenten ist es besonders wichtig, kognitive Modelle und Lernmodelle (siehe auch Kap. 4.1 in den Entwicklungsprozess einfließen zu lassen. Diese Modelle beschreiben, wie Schüler/innen Wissen repräsentieren und Kompetenzen in einer bestimmten Domäne erwerben (Pellegrino & Chudowsky, 2003). Vor allem der Unterschied zwischen Personen, die bereits Wissen in einer Domäne erworben haben (Experten), und solche, denen das Wissen noch fehlt und die sich daher noch im Lernprozess befinden (Novizen), kann Aufschluss darüber geben, wie sich Kompetenzen entwickeln. Man könnte in Hinblick auf die Testkonstruktion fragen,

- was Experten von Novizen in einem bestimmten Fach unterscheidet?
- wie in diesem Fach Novizen zu Experten werden, wie erlangen sie Kompetenzen?
- welche Strategien Lernende verwenden, um sich die Inhalte in dem jeweiligen Fach anzueignen?

Expertenwissen ist nicht nur gekennzeichnet durch eine große Menge an gespeicherten Gedächtnisinhalten in einer bestimmten Domäne, sondern auch durch eine bessere Verfügbarkeit und Abrufbarkeit dieser Inhalte. Daraus lässt sich ableiten, dass Inhalte (oder Fakten) in der Klasse hinsichtlich ihrer Anwendbarkeit unterrichtet werden sollten, bzw. dass Kompetenzmessungen berücksichtigen sollten, inwieweit Schüler/innen erkennen, wann, wo und wie sie ihr Wissen anwenden können (National Research Council, 1999; Pellegrino & Chudowsky, 2003).

Der Hauptunterschied zwischen Experten und Novizen besteht darin, dass Experten neue Situationen als minimale Abweichungen von bisher Gelerntem wahrnehmen. Dadurch können sie *starke* Strategien (siehe Kap. 4.2) aus ihrem Wissen auch auf neue Situationen transferieren. Kinder sollten daher verstehen lernen, was an einem Problem anders ist und auch, was in Relation zu anderen Fragestellungen gleich ist.

Um in einer Leistungsmessung zwischen Experten und Novizen unterscheiden zu können, müssen Aufgaben entwickelt werden, die in ihrem Schwierigkeitsgrad variieren. Schülergruppen, die zum Beispiel die Bildungsstandards nur teilweise erreichen, benötigen noch erhöhten Förderbedarf, da sie schwierigere Aufgaben der höheren Kompetenzstufen (Bildungsstandards erreicht/übertroffen) nur mit geringer Wahrscheinlichkeit lösen können.

3.1 Der Konstruktrahmen

Theorie und Empirie lassen nicht immer die gleichen Schlüsse zu (Gorin, 2006). Das theoretische Konstrukt legt vorab fest, welches Wissen, welche Fähigkeiten und Fertigkeiten in einer bestimmten Domäne verlangt sind. Das heißt, man versucht zu bestimmen, welche Schlüsse man über Schülerleistungen machen möchte. Dies muss nicht zwingend mit der Empirie übereinstimmen, denn es können auch für das Konstrukt irrelevante Prozesse bei der Beantwortung eines Items eine Rolle spielen.

Können mit einem Item die intendierten kognitiven Prozesse eines Konstrukts erfasst werden, so erreicht man eine hohe Konstruktvalidität, der Testwert kann somit valide interpretiert werden. Man schließt allerdings meist nicht von einem Item

auf die kognitiven Prozesse, sondern von vielen Itemantworten auf die darunterliegenden Prozesse. Dadurch wird es notwendig, sowohl ein kognitives Modell für das intendierte Konstrukt als auch eines für das aus der Beobachtung hervorgegangene Konstrukt zu erarbeiten.

Letzteres kann nur auf Itemebene gemacht werden und oft ist es schwierig genau zu sagen, welche Prozesse für die Beantwortung eines Items wichtig sind. Eine Möglichkeit, um sich so ein Modell zu erarbeiten, wäre, Itemmerkmale zu ermitteln, die den Komplexitätsgrad beeinflussen.

3.2 Itemmodelle

Wie erkennt man, dass ein Item schwieriger ist als ein anderes? Zwei Dinge sind hier wesentlich: komplexere Items benötigen generell eine längere Bearbeitungszeit (was allerdings meist nicht erfasst werden kann) und werden von weniger Schülerinnen und Schülern gelöst. Die Schwierigkeit von Items kann statistisch zum Beispiel durch die prozentuale Häufigkeit richtiger Antworten eines Items ausgedrückt werden. Löst das Item ein geringer Prozentsatz an Schülerinnen und Schülern, handelt es sich wahrscheinlich um ein schweres Item. Warum Schüler/innen eine bestimmte Aufgabe weniger häufig lösen, lässt sich allerdings nicht klar sagen. Dies kann meist nur auf theoretischer Ebene erfolgen. Das heißt, es muss ein Itemmodell erstellt werden, das am besten die Itemkennwerte (i. e. Itemschwierigkeit und Reaktionszeit) statistisch beschreiben kann.

Beispiel: Ein bestimmtes Leseitem wird nur von einer geringen Prozentzahl an Lernenden korrekt gelöst. Das Item besitzt sehr lange und verschachtelte Sätze. Eine Hypothese wäre nun, dass die Satzlänge (Merkmal 1) und die Anzahl an Nebensätzen (Merkmal 2) die Schwierigkeit des Items beeinflussen. In einem statistischen Modell könnte man nun diese beiden Merkmale verwenden, um generell bei Items die Schwierigkeit vorherzusagen. Ist das statistische Modell korrekt, so könnte man bei Items mit langen Sätzen schon vorab sagen, dass diese schwieriger sind als Items mit kurzen Sätzen. Das Gleiche würde für das zweite Merkmal gelten.

Was macht Items schwierig? Itemmodelle versuchen den Komplexitätsgrad von Items zu erklären, indem als Erstes Prozesse aufgeführt werden, die sequenziell für die Beantwortung eines Items notwendig sind, und als Nächstes beobachtbare Itemmerkmale definiert werden, welche für jedes Item kodiert werden².

Im Falle eines Lesetests könnten zuerst Prozesse definiert werden, die in zeitlicher Abfolge zur Aufgabenbewältigung wichtig sind. Ein erster Prozess wäre somit das einfache Enkodieren von Wörtern und Sätzen. Itemmerkmale, welche diese Prozesse in ihrer Schwierigkeit mitbestimmen, wären hier das verwendete Vokabular, die Aussagendichte etc. Als nächster Prozess wäre das Integrieren von Informationen relevant, um Kohärenz zu erreichen. Ein Itemmerkmal könnte hier die Länge einer bestimmten Passage sein.

Beispiel: In einer Konstruktbeschreibung zum Leseverständnis gibt es folgenden Punkt: *Die Schüler und Schülerinnen können zentrale Inhalte eines Textes verstehen.* Es werden nun Items entwickelt, die genau diese Kompetenz erfassen sollen. Dies ist allerdings schwierig, da keinerlei Angaben gemacht werden, was genau zentrale Inhalte sind. In einem Itemmodell kann festgelegt werden, welche (kognitiven)

² Durch die Kodierung können dann die Itemmerkmale statistisch zur Vorhersage der Itemschwierigkeit verwendet werden.

Prozesse für diese Kompetenz wichtig sind und welche Merkmale von Items diese Prozesse erfassen.

Dieses Modell ist wesentlich detaillierter als die vorangegangene Konstruktbeschreibung. Zum Beispiel gibt das Merkmal *Satzlänge* an, wie lang die Sätze des Textes sind, das Merkmal *Typ an Information* besagt, ob die zu suchende Information im Text konkret oder abstrakt vorliegt (siehe Kap. 3.3), also eher einfach zu finden ist oder schwieriger. Für das Verständnis von *zentralen Inhalten* sind möglicherweise Itemmerkmale wichtig, die Prozesse erfassen, wie die Information aus dem Text extrahiert werden kann. Findet man die richtige Information in einem Absatz (lokales Verständnis) oder muss man die Information durch einen Integrationsprozess erschließen, der das Zusammensuchen von vielen Einzelinformationen beinhaltet (globales Verständnis). Letzteres misst wohl am ehesten die im Konstrukt definierte Kompetenz *zentrale Inhalte verstehen*. Das Itemmodell beschreibt somit genauer, welche Itemmerkmale beinhaltet sein sollten, um einen Aspekt eines Konstrukts (oder eine Kompetenz) messen zu können. In Leistungstests möchte man besonders Schülergruppen, die diese Kompetenz besitzen, von Schülergruppen, die diese Kompetenz nicht besitzen, unterscheiden können. Man konstruiert demnach Items, welche das Itemmerkmal für die jeweilige Kompetenz beinhalten (und solche, die das nicht oder weniger tun) und auch in der Schwierigkeit variieren (um auch Leistungsunterschiede feststellen zu können). Das Variieren der Itemmerkmale und der Schwierigkeit von Aufgaben ist somit zentral für die Messung von Schülerkompetenzen.

Ein bisher nicht beschriebenes Problem in Zusammenhang mit Itemschwierigkeit kann allerdings auftreten, wenn die Schwierigkeit durch das Antwortformat verzerrt ist: wenn z. B. zwei Aufgaben theoretisch gleich schwierig wären und womöglich auch das Gleiche testen sollten (die gleichen Kompetenzen), durch das Format oder auch durch die Formulierung der Frage aber unterschiedlich häufig gelöst würden. Das Antwortformat ist somit auch ein wichtiges Itemmerkmal, welches durchaus schwierigkeitsbestimmend sein kann, allerdings nicht in einem theoretischen Konstrukt festgehalten wird.

3.3 Schwierigkeitsbestimmende Merkmale von Items

Schwierigkeitsbestimmende Merkmale von Testitems sind wesentlich für die Entwicklung von Kompetenzstufenmodellen und die Erstellung von Items (Böhme, Robitzsch & Buse, 2010; Bremerich-Vos & Böhme, 2009; Buck & Tatsuoka, 1998; Freedle & Kostin, 1993; Grotjahn, 2000, 2001; Gorin, 2006; Gorin & Embretson, 2006; Isaac & Hochweber, 2011; Kirsch, Jungeblut & Mosentahl, 1998; Leong, 2006)³. So sollten Itementwickler/innen versuchen möglichst ansprechende Items zu entwickeln, die vorwiegend höhere kognitive Verarbeitungsprozesse messen (The Donath Group, 2005). Diese Prozesse umfassen Verstehen, Problemlösen, kritisches Denken, Analyse, Synthese, Evaluation und Interpretation im Gegensatz zum Abrufen von deklarativem Wissen (vgl. Kap. 4.1, Kap. 4.2 und Abb. 3). Nur Schülerinnen und Schüler, die eine bestimmte Kompetenz oder Kompetenzen aufweisen, sollen das entsprechende Item lösen können. Das Nachfolgende bezieht sich vermehrt auf die Lesekompetenz, die erwähnten Itemmerkmale können allerdings auch für andere Kompetenzbereiche und Fächer adaptiert werden.

³ Empirische Untersuchungen zu schwierigkeitsbestimmenden Merkmalen bedienen sich regressionsanalytischer Methodiken, bei denen die Schwierigkeit von Items als gewichtete Summe von Aufgabenmerkmalen modelliert wird. In anderen Untersuchungen findet man auch die Verwendung von linear-logistischen Testmodellen (LLTM), diese bieten den Vorteil, in einem einstufigen Verfahren die Merkmale von Items mitzuberücksichtigen. Sie lassen allerdings keinen Fehlerterm zu, was in Regressionsanalysen gegeben ist (Varianzaufklärung).

Speziell für das Lesen sind reine Oberflächenmerkmale, wie Anzahl verschiedener Wörter, durchschnittliche Anzahl von Silben pro Wort, Anzahl der Sätze etc. für die Vorhersage von Leseverständnis weniger sinnvoll⁴. Inhalte von Texten sowie Personenmerkmale (Vorwissen, Motivation von Lernenden etc.) bleiben hier völlig unberücksichtigt (Grotjahn, 2000). So ist es zum Beispiel wichtig anregende Texte zu verfassen, in denen die inhaltlichen Beziehungen nicht sofort ersichtlich sind, der Leser sich diese Zusammenhänge erarbeiten muss. Durch die erhöhte kognitive Anstrengung wird die Motivation positiv beeinflusst. Motivation und auch Vorwissen spielen daher eine wichtige Rolle für das Textverständnis. Es sei jedoch erwähnt, dass in Schülerleistungstudien motivationale Aspekte größtenteils unberücksichtigt bleiben bzw. selten direkt erfasst werden. Ist ein Text allerdings wenig motivierend, so kann es sein, dass das Item gar nicht bearbeitet wird (siehe auch Kap. 4.1.3).

Im Wesentlichen ist bei Testaufgaben zu hinterfragen, wie die notwendige Information erschlossen werden kann (Kirsch et al., 1998). Dies trifft nicht nur auf Leseaufgaben zu, sondern kann verallgemeinert auch für Aufgaben aus anderen Fächern übernommen werden (z. B. Mathematik)⁵. Teilt man eine Aufgabe in zwei Teile, so kann man die gegebene Information von der erfragten Information trennen. Determinanten für die Aufgabenschwierigkeit sind nun jene geforderten Prozesse oder Strategien welche diese beiden Informationsquellen miteinander verknüpfen sollen, um die notwendige Information zu erschließen.

Aufgabencharakteristika gliedern sich laut Kirsch et al. (1998) in drei wesentliche Bereiche: *Inhalt/Kontext*, *Materialien/Text* und *Prozesse/Strategien* (siehe Abb. 1). Die ersten beiden Kategorien bleiben hier jedoch unberücksichtigt, obwohl diese durchaus den Komplexitätsgrad von Items beeinflussen können. Kontext bezieht sich zum Beispiel auf das Thema des Textes (z. B. Thema Familie), Materialien bezeichnet, ob in dem Text Grafiken, Tabellen etc. vorkommen. Wichtiger für die Itementwicklung sind allerdings die geforderten kognitiven Prozesse in der dritten Kategorie, auf diese soll weiter unten genauer eingegangen werden.

Abbildung 2 versucht ein erweitertes Modell, basierend auf verschiedenen Arbeiten (z. B. Bremerich-Vos & Böhme, 2009; Kirsch et al., 1998), darzustellen. Aufgabenmerkmale können demnach in vier verschiedene Kategorien eingeteilt werden. Oberflächenmerkmale bezeichnen größtenteils Merkmale quantitativer Natur (wie lang sind die Wörter, Sätze, Absätze etc.) als auch direkt ersichtliche Itemmerkmale (wie Itemformate). Die Plausibilität von Distraktoren ist bei Kirsch et al. (1998) ein separates Itemmerkmal, man könnte es allerdings auch als nicht reines Oberflächenmerkmal bezeichnen, das auch schon in Richtung Inhaltsmerkmal tendiert. Dieses Merkmal beschreibt, wie nahe die einzelnen Falschantworten an der richtigen Antwort liegen: Je näher, umso schwieriger das Item, je unplausibler (wenig nahe), umso einfacher das Item.

Die nächste Kategorie bezeichnet sowohl inhaltliche Merkmale als auch die Interaktion Person mit dem Text. Metaphern (inhaltliches Merkmal) erschweren die Texte meist. Wie bereits erwähnt, interagieren Faktoren wie Motivation und Vorwissen meist mit Items und Itemmerkmalen (dies ist natürlich kein inhärentes Itemmerkmal, sondern führt indirekt zu höheren Itemschwierigkeiten). Haben Lernende ein allgemein höheres Motivationsniveau, so werden sie auch Items lösen, die textlich anspruchsvoller sind. Andererseits können sehr komplexe Texte bei Personen mit niedrigem Motivationsniveau dazu führen, diese erst gar nicht zu bearbeiten. Zusätzlich kann

⁴ Die Lesbarkeitsforschung leistet hier dennoch einen wesentlichen Beitrag, da zum Beispiel Lesegenauigkeit und Lesegeschwindigkeit durchaus mit Oberflächenmerkmalen korrelieren.

⁵ Merkmale für Hörverständnis findet man bei Buck und Tatsuoaka (1998), für Leseverständnis bei Freedle und Kostin (1993); Kirsch et al. (1998) zeigen ein Merkmalsystem sowohl für Lesen als auch für Mathematik.

ein Text, je nachdem wie ansprechend dieser ist, das individuelle Motivationsniveau erhöhen oder verringern. Die Motivation wird zwar im Leistungstest meist nicht berücksichtigt, es soll allerdings erwähnt werden, dass deren Grad durchaus die Lösungswahrscheinlichkeit eines Items beeinflussen kann. Vorwissen im Sinne von Alltags- oder Weltwissen sollte nicht ausschließlich für die Beantwortung eines Items ausreichen, lässt sich allerdings nicht immer vermeiden (bzw. ist manchmal auch gewünscht, siehe PISA-Studie). Vorwissen kann auch begriffliches Wissen miteinschließen, was in manchen Kompetenzbereichen natürlich verlangt werden kann.

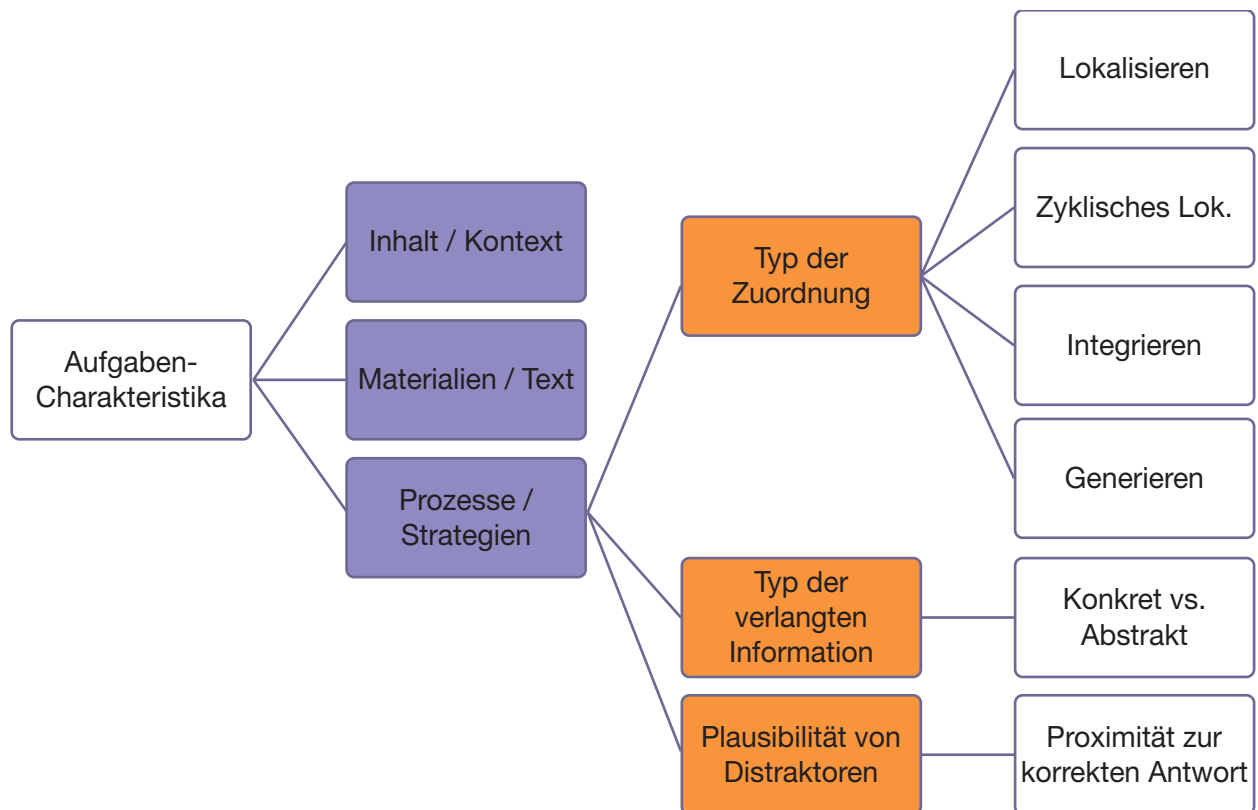


Abbildung 1: Vereinfachtes Modell von schwierigkeitsbestimmenden Aufgabenmerkmalen nach Kirsch et al. (1998).

Die erfragte Information kann nun

- in *konkreter* Form vorliegen (Frage nach Objekten, Gegenständen, Orten, Namen etc.);
- in *abstrakter* Form vorliegen (Ursachen, Folgen, Zwecke etc.).

Es ist wichtig zu erwähnen, dass diese beiden Merkmale keineswegs diskret vorliegen müssen, sondern es ist vielmehr davon auszugehen, dass sie auf einem Kontinuum zu finden sind. Das heißt, die Information kann mehr oder weniger abstrakt oder konkret vorliegen, die Information liegt aber selten entweder abstrakt oder konkret vor.

Als Nächstes stellt man sich die Frage, wie die erfragte Information mit der vorhandenen, gegebenen Information verknüpft wird. Die Verknüpfung der Aufgabe mit z. B. Textinformation (Leseaufgaben) erfolgt durch vier wesentliche Prozesse (Kirsch et al., 1998, Abb. 1):

1. **Einfaches Lokalisieren** bezeichnet einen Prozess, bei dem die Information explizit vorliegt und dadurch auch leicht zu erschließen ist. Ein oder mehrere Informationsmerkmale aus der Frage werden nahezu identisch oder synonym den Informationsmerkmalen aus dem Text zugeordnet. Dies wird auch mit Bezeichnungs- oder Bedeutungsgleichheit umschrieben.
2. **Zyklisches Lokalisieren** ist ein sehr ähnlicher Prozess wie der erstgenannte, es sind hier allerdings mehrere Durchläufe erforderlich, um die notwendige Information zu erhalten, eine Serie an Informationszuordnungen muss hier gemacht werden.
3. **Integrieren** schließt auch die Verknüpfung von mehreren Informationen auf Basis von Ähnlichkeiten, Differenzen, Ursache-Wirkungs-Relationen etc. mit ein.
4. **Generieren**, der komplexeste Prozess, und somit der Prozess der wesentlich die Schwierigkeit von Items beeinflusst, bindet auch Vorwissen ein, um die notwendige Information zu erfassen. Vorwissen bezieht sich allerdings nicht (nur) auf Weltwissen, sondern umfasst vielmehr domänenspezifisches Vorwissen (z. B. Sprachwissen bei Leseverständnis). Nicht nur textbasierte, sondern auch vorwissensbasierte Schlüsse müssen gezogen werden. Ist etwa die Aufgabe, in einem Branchenverzeichnis eine Person zu finden bei der man Brillen kaufen möchte, so ist es wichtig zu wissen, dass auch Optiker Brillen verkaufen (Bremerich-Vos & Böhme, 2009).

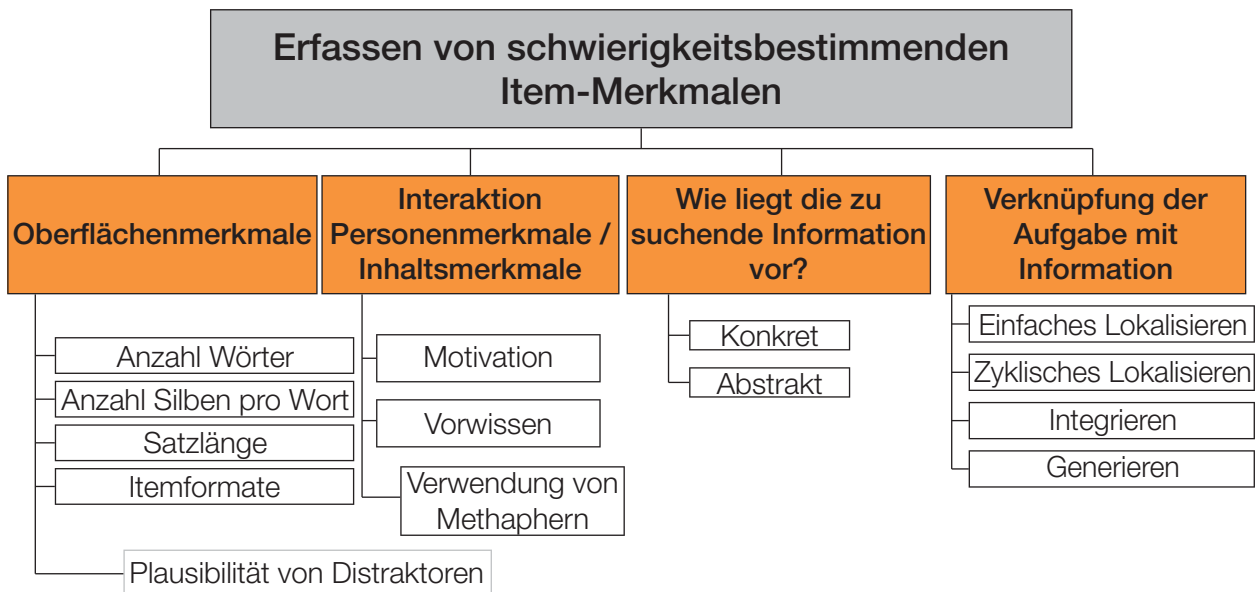


Abbildung 2: Modell von schwierigkeitsbestimmenden Aufgabenmerkmalen.

4 Exkurs

4.1 Lernzieltaxonomien

Es soll hier vor allem auf Denkprozesse eingegangen werden, die für die Bearbeitung von Aufgaben, aber auch für die Aneignung von Wissen bei Schülerinnen und Schülern von Relevanz sind. Als grundlegendes Rahmenwerk zum Verständnis von Denkprozessen beim Lernen dient die Taxonomie von Bloom (Bloom, Engelhart, Furst & Hill, 1956). Neuere Überarbeitungen dieser Taxonomie weisen auf mögliche Probleme hin, die in der ursprünglichen Arbeit unzureichend berücksichtigt worden sind (Marzano & Kendall, 2007). Die einzelnen Facetten, die in diesen Modellen beschrieben sind, liefern ein sehr gutes Verständnis für die kognitiven Prozesse während des Lernens von Neuem bzw. auch Hintergründe zu Denkabläufen bei der Bearbeitung von Aufgaben in Tests oder Schülerleistungsstudien (Cheng, 2006; Dimitrov & Raykov, 2003; Wendt & Harnes, 2009)⁶.

Für die Itementwicklung sind diese Modelle dahingehend wichtig, dass sie Aufschluss über mögliche Komplexitätsstufen geben können. Die Originalarbeit von Bloom beschreibt zum Beispiel eine hierarchische Struktur des Lernens, bei der die Komplexität mit jeder Stufe zunimmt. Diese Hierarchie wurde in späteren Modellen kritisch hinterfragt. Wichtige Schlussfolgerungen hinsichtlich kognitiver Einflussgrößen, die sich für die Generierung von unterschiedlich schwierigen Items ableiten lassen, werden hier kurz erläutert.

4.1.1 Blooms Taxonomie

Generell bezeichnet eine Taxonomie ein Kategorisierungssystem. In der Botanik findet man zum Beispiel Taxonomien zur Klassifizierung unterschiedlicher Pflanzenarten, Pflanzenmerkmale können somit verschiedenen Kategorien (Blattgröße, Farbe etc.) zugeordnet werden. Ähnlich geschieht dies bei der Lernzieltaxonomie von Bloom. Hier werden Lernziele in verschiedene Kategorien eingeteilt. Die Bloom'sche Taxonomie geht von sechs Kategorien aus⁷: Wissen, Verständnis, Anwendung, Analyse, Synthese und Evaluation (siehe Abb. 3).

Die Taxonomie ist hierarchisch nach Komplexitätsgrad strukturiert, die Komplexität nimmt somit mit aufsteigender Reihenfolge zu. Zusätzlich wurde angenommen, dass dieses Kategoriensystem auch kumulativ sei, das heißt, komplexere Verhaltensweisen bauen auf einfacheren auf (Mietzel, 2007).

Die unterste Domäne Wissen bezeichnet Prozesse, die mit Erinnerungsleistung zu tun haben. Verständnis als zweite Domäne beschreibt Fähigkeiten der Informationsaufnahme bzw. der Verarbeitung. Die Domäne der Anwendung bezieht die Abstraktionsfähigkeit ein, die Analyse verbindet die Domäne Verständnis und Anwendung. Unter Synthese wird die Fähigkeit zur Generierung von neuen Wissensstrukturen verstanden. Der höchste Level, Evaluation, verlangt nach gesteigerter bewusster Kontrolle da hier Entscheidungen über den Wert von Wissen getroffen werden.

⁶ Es existieren zurzeit allerdings noch keine klaren empirischen Befunde aus Studien, welche den Einsatz von Lernziel-Taxonomien oder kognitiven Modellen bei der Itemgenerierung untersucht haben (The Donath Group, 2005).

⁷ Diese werden wiederum in Subkategorien unterteilt, auf die hier nicht eingegangen wird.

Marzano und Kendall (2007) geben eine Übersicht über die Hauptkritikpunkte zum Bloom'schen Modell. Vor allem der hierarchische Ansatz hat sich empirisch nicht bestätigen lassen, eine höhere Komplexität mit zunehmender Stufe kann somit nur schwer angenommen werden.

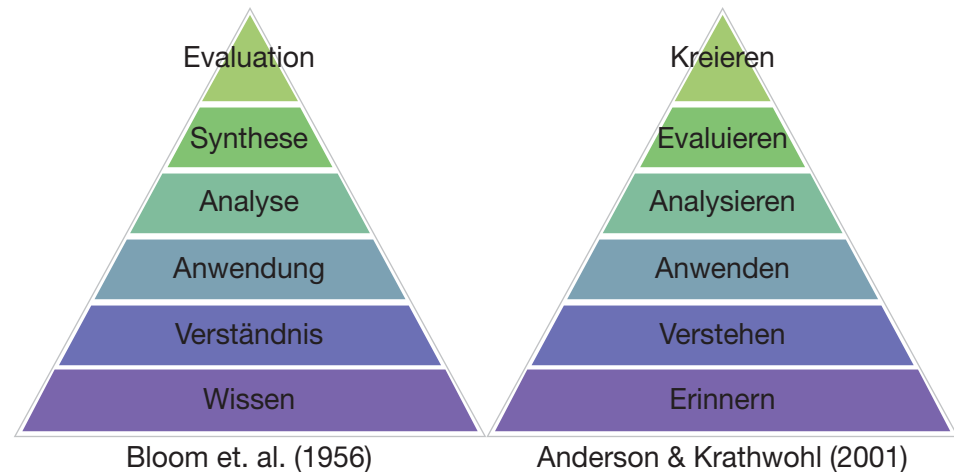


Abbildung 3: Lernzieltaxonomien.

4.1.2 Überarbeitung der Bloom'schen Taxonomie

Die populärste Überarbeitung des Original-Modells findet man bei Anderson und Krathwohl (2001). Das eindimensionale Modell von Bloom wird in ein zweidimensionales Modell mit den Dimensionen Wissen (knowledge) und Kognitive Verarbeitung (cognitive process) gegliedert. Die Wissensdimension umfasst vier Subkategorien, die anhand eines Beispiels aus Anderson und Krathwohl (2001) kurz beschrieben werden sollen:

■ **Faktenwissen:**

Diese Kategorie beschreibt das Grund- oder Basiswissen, das zur Problemlösung notwendig ist. Am Beispiel eines bestimmten literarischen Stücks könnte ein Unterrichtender etwa an Details wie Hauptpersonen des Stücks, Textpassagen, Dialoge, etc. interessiert sein.

■ **Konzeptuelles oder begriffliches Wissen:**

Wird vermehrt Wert darauf gelegt, dass ein/e Schüler/in die Konzepte wie Ironie, Ambition etc. eines Stücks erlernen soll, handelt es sich um das Erfassen von konzeptuellem Wissen. Ähnliches wäre bei Mathematikaufgaben das Verständnis für bestimmte Theorien und Modelle (z. B.: Lehrsatz von Pythagoras, Ohm's Gesetz, etc.). Begriffe und Regeln werden bereits nicht nur als Fakten, sondern abstrahiert wahrgenommen.

■ **Prozedurales Wissen:**

In der Kognitiven Psychologie unterscheidet man zwischen deklarativem und prozeduralem Gedächtnis. Ersteres beschreibt das Faktenwissen, zweiteres beschreibt das Gedächtnis über das Wissen, *wie Dinge getan werden* (Gerrig & Zimbardo, 2008). Diese Wissensdomäne beschreibt demnach, wie bestimmte Abläufe funktionieren. Im Beispiel des literarischen Stücks, das Schülerinnen und Schülern vermittelt wird, könnte hier besonders Wert darauf gelegt werden, wie sie allgemein an die Interpretation von literarischen Werken herangehen können. Es

können verschiedene Aspekte, die zur Interpretation wichtig sind, vermittelt werden (z. B. Charaktere, Aufbau, kultureller Kontext etc.). Anhand dieser Liste können auch zukünftige Stücke interpretiert werden.

■ **Metakognitives Wissen:**

Metakognitives Wissen beinhaltet die Reflexion über das erlernte Wissen. In Bezug auf das Literaturwerk könnte man Schülerinnen und Schüler instruieren, über Probleme zu reflektieren, die sie in der Anwendung von bestimmten Prozeduren hinsichtlich der Interpretation des Stücks haben/hatten. Aus diesen Problemen können sie wiederum Dinge über dieses und auch andere Stücke verstehen, sie lernen auch, sich mit dem Stück persönlich auseinanderzusetzen, ihre eigenen Stärken und Schwächen zu erkennen und Dinge auch über die eigene Person in Erfahrung zu bringen.

Die kognitiven Prozesse können wie folgt beschrieben werden (Anderson & Krathwohl, 2001; Krathwohl, 2002):

- **Erinnern** bezieht sich auf den reinen Abruf von Wissen aus dem Langzeitgedächtnis.
- **Verstehen** ist ein breitgefächertes Prozess, bei dem neues Wissen mit bereits Gelerntem in Verbindung gebracht wird. Lernende können Gelerntes interpretieren, exemplifizieren, klassifizieren, zusammenfassen, abstrahieren, vergleichen und erklären. Das neu erarbeitete Wissen, so Mietzel (2007), muss mit Sinn gefüllt werden, dies geschieht durch den Vergleich mit bereits erworbenem Wissen. Die Schüler/innen können also Vergleiche herstellen oder Beispiele geben.
- **Anwenden** beinhaltet die Prozesse des Ausführens und Implementierens. Schüler/innen können erlernte Routinen ausführen, wenn sie mit vertrauten Aufgaben konfrontiert werden und können auch relevante Routinen auswählen, die sie auf neue, ungewohnte Situationen übertragen können. Sie demonstrieren damit Verständnis für gelernte Sachverhalte.
- **Analysieren** beschreibt einen Prozess, bei dem das Material in seine Bestandteile aufgeteilt wird (Differenzieren) sowie die Beziehungen zwischen den einzelnen Teilen (Organisieren) und der übergeordneten Struktur (Attribuieren) ersichtlich wird. Schüler/innen können zum Beispiel die Hauptidee eines Textes von Nebenaussagen, relevante Aussagen von nichtrelevanten unterscheiden. Sie können Evidenz für die Hauptaussagen des Autors finden, Fakten von Meinungen trennen und Konklusionen hinsichtlich ihrer Kernaussagen erfassen.
- **Evaluieren** heißt, es werden Urteile basierend auf Kriterien und Standards gemacht. Urteile können über die Effizienz oder Qualität verschiedenster Methoden und Probleme getroffen werden. Werden zum Beispiel Fragen gestellt, wie sinnvoll ein Vorgehen ist, ob dieser Prozess oder diese/jene Methode gut und effizient ist etc., dann handelt es sich um dem Prozess des Evaluierens oder auch Bewertens.
- **Kreieren** bezeichnet den Prozess, bei dem Dinge zusammengesetzt werden, um ein kohärentes Ganzes zu ergeben. Bisher Gelerntes und bisherige Lernerfahrungen werden verwendet, um Neues zu erschaffen. Verschiedenste Gebiete müssen miteinbezogen werden, um ein neues Produkt zu generieren. Bevor ein neues Produkt entsteht, muss als Erstes das Problem analysiert werden, um zu einer genauen Repräsentation des Problems zu kommen (Generieren). Neue Lösungswege werden in einem kreativen Prozess bereits erdacht. Als Nächstes wird ein Plan zur Problemlösung erstellt (Planen), welcher anschließend ausgeführt wird (Produzieren).

Alle diese Punkte können verwendet werden, um Lernziele zu definieren und daraus Instruktionen für Schüler/innen zu entwickeln.

4.1.3 Eine neue Taxonomie (NT)

Die hierarchische Form der Bloom'schen Taxonomie und die damit verbundene Definition von Aufgabenschwierigkeit konnte empirisch nicht gezeigt werden und war daher für einige Autoren Anlass zur Entwicklung einer neuen Lernziel-Taxonomie. Marzano und Kendall (2007) gehen von zwei wesentlichen Merkmalen aus, die die Komplexität von kognitiven Vorgängen bestimmen. Zum einen existiert laut den Autoren eine inhärente Komplexität innerhalb eines bestimmten kognitiven Prozesses, determiniert durch die einzelnen Schritte, die zur erfolgreichen Durchführung notwendig sind. Zum Anderen bestimmt der Vertrautheitsgrad mit einem bestimmten Prozess dessen Schwierigkeit. Das Schnüren von Schubhändlern ist zum Beispiel ein komplexer Prozess mit bestimmten Zwischenschritten, der allerdings durch die häufige Anwendung uns sehr vertraut und daher wenig schwierig erscheint. Die Autoren gehen somit weg von einer hierarchischen Definition von Komplexität und ordnen mentale Prozesse hinsichtlich des Grades an Kontrolle.

Im Gegensatz zur Bloom'schen Taxonomie erhebt die NT den Anspruch einer Theorie, es können somit Vorhersagen über Verhalten getroffen und empirisch überprüft werden. Im Wesentlichen beschreibt die NT einen hierarchischen Informationsfluss bei der Bearbeitung von bestimmten Aufgaben (siehe Abb. 4).

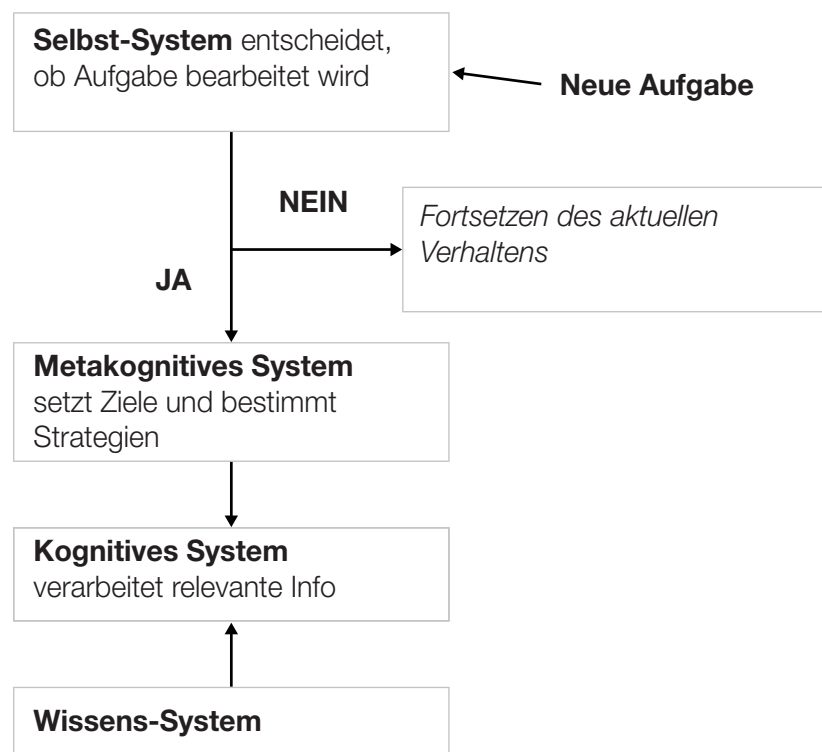


Abbildung 4: Modell der neuen Taxonomie.

Das *Selbstsystem* beinhaltet Einstellungen, Ziele, Selbstkonzepte und motivationale Faktoren. Die Entscheidung, ob man sich einer Aufgabe hingibt, wird hier getroffen. Sind Aufgaben wenig motivierend oder zu schwierig formuliert, wird ein Scheitern oder Nichtbearbeiten bereits hier bestimmt. Entscheidet man sich für eine Aufgabe, legt das *Metakognitive System* die Ziele zur Bewältigung und zum erfolgreichen Lernen der neuen Materie fest. Das *kognitive System* dient schließlich der eigentlichen Informationsverarbeitung, analytische Operationen werden hier getätigt. Zusätzlich

nimmt die NT noch ein separates *Wissens-System* an, welches in ständigem Kontakt mit dem kognitiven System steht. Lernziele werden somit von bereits erworbenen Wissensstrukturen mitbestimmt.

4.2 Gedächtnissysteme

Für den Erwerb von Wissen sind unterschiedliche Faktoren wichtig. Aus der kognitiven Lernforschung sind im Wesentlichen drei große Themenbereiche von besonderem Interesse: Gedächtnis, Aufmerksamkeit und Motivation. Für die folgenden Erläuterungen zu den bestimmten Modellen wird hier besonders auf den Gedächtnispekt eingegangen, da vor allem dieser im Zentrum steht (siehe Gerrig & Zimbardo, 2008; Tulving & Craik, 2000).

Die Gedächtnisinhalte können in domänenspezifisches und domänenübergreifendes Wissen eingeteilt werden. Hinsichtlich Problemlösestrategien repräsentieren spezifische Wissensinhalte starke und allgemeine Wissensinhalte schwache Methoden, um zum Beispiel eine Aufgabe zu lösen. Vor allem bei neuartigen Problemen können meist nur domänenübergreifende, schwache Methoden zur Problemlösung herangezogen werden, wie zum Beispiel die Versuch-Irrtum- Methode (engl.: Trial-Error). Für die effiziente Problemlösung werden natürlich starke Methoden bevorzugt. Können also Schüler/innen Probleme als neuartige Situationen erkennen, welche nur Abweichungen zu bisher gelernten Situationen darstellen, können sie leichter starke Methoden anwenden.

Die Gedächtnissysteme sind in Abbildung 5 dargestellt. Das Arbeitsgedächtnis (AG) hat eine beschränkte Kapazität und Inhalte werden nur für kurze Dauer abgespeichert. Gedächtnisinhalte können in das Langzeitgedächtnis (LZG) übertragen bzw. auch von dort abgerufen und in den Arbeitsspeicher transferiert werden. Sind Gedächtnispfade gefestigt, spricht man von einer Konsolidierung, die Inhalte werden im LZG abgespeichert. Neuere Forschung konnte zeigen, dass Gedächtnisinhalte, die bereits im LZG gespeichert wurden und für eine bestimmte Aufgabe wieder in den Arbeitsspeicher gelangen, also uns bewusst sind, in einem sehr fragilem Zustand sind. In dieser Phase können Inhalte einem Update unterzogen werden, man spricht hier von Rekonsolidierung. Unter anderem können diese Inhalte auch ganz vergessen werden.

Das LZG umfasst zwei weitere Subsysteme: das deklarative und das prozedurale Gedächtnis. Wie schon die Namen vermuten lassen, umfasst das deklarative Gedächtnis Inhalte, die benannt (engl.: to declare) werden können, wie z. B. *der Himmel ist blau* oder *der Papst lebt in Rom*. Dieses System kann weiters in ein semantisches Gedächtnis (Gedächtnis über Fakten, Wien ist die Hauptstadt von Österreich) und ein episodisches Gedächtnis unterteilt werden. Das letztere beinhaltet Wissen über persönliche Ereignisse, wie z. B. *der Kuchen auf der Hochzeit meiner Schwester war eine Sachertorte*.

Wie bereits erwähnt, ist das Arbeitsgedächtnis in seiner Kapazität stark begrenzt, warum dies der Fall ist, ist weiterhin unklar (Freunberger, Werkle-Bergner, Griesmayr, Lindenberger & Klimesch, 2011). Im Grunde können im Arbeitsgedächtnis nur 7 ± 2 Gedächtniseinheiten (z. B. Buchstaben oder Zahlen) aufgenommen werden, für das visuell-räumliche Arbeitsgedächtnis ist dies sogar auf 4 ± 1 Einheiten begrenzt. In gehirnephysiologischen Untersuchungen konnte gezeigt werden, dass vor allem die Unterdrückung von irrelevanter Information zu verbesserter Gedächtnisleistung führt. Personen, die nichtrelevante Information besser ausblenden können, zeigen bessere Gedächtnisleistung als Personen, die auch irrelevante Informationen verarbeiten.

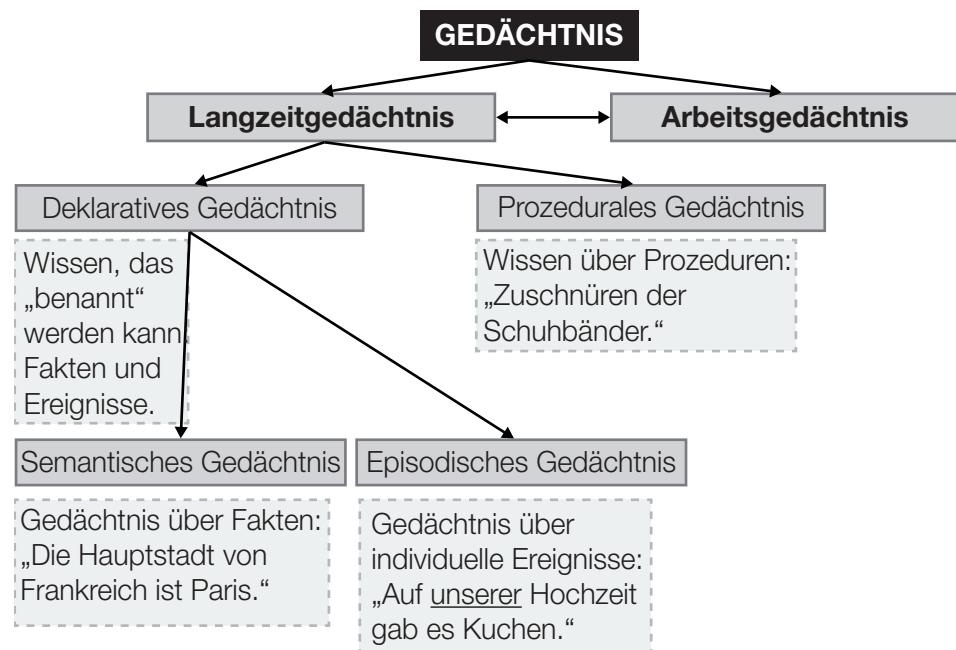


Abbildung 5: Das Gedächtnissystem.

Hinsichtlich Itemwriting sollten die Besonderheiten des Gedächtnissystems berücksichtigt werden. Vor allem das Arbeitsgedächtnis wird unterschiedlich beansprucht, wenn es zum Beispiel um gesprochene (z. B. Höritems) gegenüber geschriebener Sprache (z. B. Leseitems) geht (Kürschner & Schnotz, 2008). Redundanz ist daher beim Sprechen (Hören) wesentlich wichtiger als bei geschriebenem Text, da das Arbeitsgedächtnis in seiner Aufnahmefähigkeit begrenzt ist und Gesprochenes vergänglich ist, das heißt, stärker im Arbeitsgedächtnis verankert werden muss, um wiedererinnert werden zu können.

Literatur

- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning teaching and assessing: a revision of bloom's taxonomy of educational objectives*. New York: Longman.
- Böhme, K., Robitzsch, A. & Buse, A.-K. (2010). Zur Abgrenzung des Hörverstehens gegen über dem Leseverstehen mit Hilfe schwierigkeitsbestimmender Merkmale bei der Entwicklung von Testaufgaben. In V. Bernius & M. Imhof (Hrsg.), *Zuhörkompetenz in Unterricht und Schule* (S. 81–104). Göttingen: Vandenhoeck & Ruprecht.
- Bloom, B. E., Engelhart, M., Furst, E. & Hill, D., W.H. Krathwohl. (1956). *Taxonomy of educational objectives: handbook i: cognitive domain*. New York: David McKay.
- Bremerich-Vos, A. & Böhme, K. (2009). Lesekompetenzdiagnostik - die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzer & O. Köller (Hrsg.), (S. 219–249). Weinheim: Beltz.
- Buck, G. & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing*, 15, 119–157.
- Cheng, L. (2006). On varying the difficulty of test items. In *Paper presented at the 32nd annual conference of the international association for educational assessment, singapore*.
- Dimitrov, D. & Raykov, T. (2003). Validation of cognitive structures: a structural equation modeling approach. *Multivariate Behavioral Research*, 38, 1–23.
- Freedle, R. & Kostin, I. (1993). The prediction of toe reading item difficulty: implications for construct validity. *Language Testing*, 10, 133–170.
- Freunberger, R., Werkle-Bergner, M., Griesmayr, B., Lindenberger, U. & Klimesch, W. (2011). Brain oscillatory correlates of working memory constraints. *Brain Research*, 1375, 93–102.
- Gerrig, R. J. & Zimbardo, P. G. (2008). *Psychologie*. München: Pearson Studium.
- Gorin, J. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21–35.
- Gorin, J. & Embretson, S. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement*, 30, 394–411.
- Grotjahn, R. (2000). Determinanten der Schwierigkeit von Leseverstehensaufgaben: Theoretische Grundlagen und Konsequenzen für die Entwicklung des TESTDAF. In S. Bolton (Hrsg.), *TESTDAF: Grundlagen für die entwicklung eines neuen sprachtests. beiträge aus einem expertenseminar* (S. 7–55). Köln: VUB Gilde.
- Grotjahn, R. (2001). Determinants of the difficulty of foreign language reading and listening comprehension tasks: Predicting task difficulty in language tests. In H. Pürschel & U. Raatz (Hrsg.), *Tests and translation. papers in memory of Christine Klein-Braley* (S. 79–102). Bochum: AKS-Verlag.

- Hartig, J. & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Berlin: Springer.
- Isaac, K. & Hochweber, J. (2011). Modellierung von Kompetenzen im Bereich Sprache und Sprachgebrauch untersuchen mit schwierigkeitsbestimmenden Aufgabenmerkmalen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 186–199.
- Kirsch, I., Jungeblut, A. & Mosentahl, P. (1998). The measurement of adult literacy. In T. Murray, I. Kirsch & L. Jenkins (Hrsg.), *Adult literacy in OECD countries: Technical report on the 1st international adult literacy survey*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Into Practice*, 41, 212–218.
- Kürschner, C. & Schnotz, W. (2008). Das Verhältnis gesprochener und geschriebener Sprache bei der Konstruktion mentaler Repräsentationen. *Psychologische Rundschau*, 59 (h), 139–149.
- Leong, S. (2006). *On varying the difficulty of test items*. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore.
- Marzano, R. J. & Kendall, J. S. (2007). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin Press.
- Mietzel, G. (2007). *Pädagogische Psychologie des Lernens und Lehrens*. Göttingen: Hogrefe.
- Mislevy, R. (2006). Cognitive psychology and educational assessment. In R. Brennan (Hrsg.), *Educational measurement*. American Council on Education.
- National Research Council. (1999). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Pellegrino, J. & Chudowsky, N. (2003). The foundations of assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1, 103–148.
- Rupp, A. A., Leucht, M. & Hartung, R. (2006). Die Kompetenzbrille aufsetzen: Verfahren zur multiplen Klassifikation von Lernenden für Kompetenzdiagnostik in Unterricht und Testung. *Unterrichtswissenschaft*, 34, 195–219.
- Rupp, A. A. & Templin, J. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6, 219–262.
- The Donath Group. (2005). *Test item writing best practices*. Autor.
- Tulving, E. & Craik, F. (Hrsg.). (2000). *The Oxford handbook of memory*. Oxford University Press.
- Wendt, A. & Harmes, J. (2009). Developing and evaluating innovative items for the nclex. *Nurse Educator*, 34, 109–113.

