

Kognitionspsychologische Aspekte eines Kompetenzmodells zur Bild-Text-Integration

Mark Ullrich, Wolfgang Schnotz, Holger Horz, Nele McElvany
Sascha Schroeder und Jürgen Baumert

Zusammenfassung. Der Projekttitel *BiTe* steht für ein Forschungsvorhaben, das die schulische Kompetenzentwicklung im Bereich der integrierten Verarbeitung von Bild- und Textinformationen untersucht. Ausgangspunkt sind dabei kognitionspsychologische Überlegungen zu den mentalen Prozessen, die dem Wissenserwerb mit Texten und Bildern zugrunde liegen. So lässt sich anhand des Modells zum integrierten Text-Bild-Verstehen von Schnotz und Bannert (2003) ableiten, dass die gemeinsame Verarbeitung von verbalen und visuellen Informationen besondere Anforderungen an die Schülerinnen und Schüler stellt. Deren Bewältigung verlangt das Vorhandensein einer grundlegenden Kompetenz, die über das *reine* Leseverstehen hinausgeht, im schulischen Kontext aber bislang nicht unterrichtet wird. Aus diesem Grund ist es ein zentrales Ziel des *BiTe*-Projekts, die Kompetenz zur integrativen Verarbeitung von Texten und Bildern fassbar zu machen, indem ein theoretisches Strukturmodell entwickelt wurde (Schnotz et al., 2010). Demzufolge beruht die integrative Verarbeitung von Texten und Bildern zu einem großen Teil auf Strukturabbildungsprozessen, die ineinander verschachtelt sind. Je nach Anzahl der beteiligten Elemente und deren Relationen lassen sich dabei drei hierarchische Kompetenzebenen unterscheiden, denen verschiedene Integrationsanforderungen zugeordnet werden können. Um die theoretischen Modellannahmen zu prüfen, wurde eine Reihe von Aufgaben entwickelt und in einer umfangreichen Studie pilotiert. Die Struktur der Kompetenz zur integrativen Verarbeitung von Texten und Bildern wurde dabei auf der Grundlage verschiedener Testmodelle evaluiert. Es zeigte sich jedoch, dass multidimensionale Modelle aufgrund hoher Korrelationen zwischen den einzelnen Skalen die Struktur nur unzureichend abbilden konnten. Schlüsselwörter: Bild-Text-Integration, Verstehen, kognitive Prozesse, Kompetenzmodelle

Cognitive psychological aspects of a competence model of picture-text-integration

Abstract. The title *BiTe* describes a research project that examines the development of competences within the domain of integrated picture and text processing in school. This study started with cognitive psychological considerations for mental processes which represent the basis for knowledge acquisition through texts and pictures. According to the model of integrated text and picture comprehension of Schnotz and Bannert (2003) the common processing of verbal and visual information puts special demands on students. In order to cope with these demands students require competencies beyond pure reading comprehension which typically have not been taught at school. Therefore, the aim of the *BiTe* project is to measure the competence to process text and picture information in an integrated way by developing a theoretical model (Schnotz et al., 2010). According to this model, the integrative processing of texts and pictures is based mainly on structural mapping processes which are linked to each other. Depending on the number of elements involved and their relationships to each other, three hierarchical levels can be differentiated to which specific integration requirements can be assigned. In order to examine the theoretical model's assumptions, a set of tasks was developed and tested in a pilot study. The structural assumption of the competence of integrating text and picture information was evaluated by comparing different test models. However, the multi-dimensional models could not fit the structure sufficiently due to high correlations between the individual scales. Key words: picture-text-integration, comprehension, cognitive processes, competence models

Die Bedeutung der Text-Bild-Integration

Die Verwendung von Bildern in Texten besitzt eine lange Tradition und lässt sich bis in die frühe Neuzeit zurückverfolgen (Schnotz, 2005). Schon damals waren Gelehrte der Ansicht, dass die Kombination von verbalen und visuel-

len Repräsentationsformaten den Wissenserwerb erleichtern und unterstützen kann. Diese Einschätzung hat sich auch empirisch als zutreffend erwiesen, wie zahlreiche Studien belegen (Carney & Levin, 2002; Levie & Lenz, 1982; Levin, Anglin & Carney, 1987). Allerdings ist es voreilig anzunehmen, dass die gemeinsame Präsentation von Text und Bild zwangsläufig zu einer besseren Lernleistung führt. Wie so oft sind die Zusammenhänge komplexer und neben den Charakteristika des Lehr-/Lernmaterials bestimmen die Merkmale des Individuums über Richtung und Ausmaß der lernförderlichen Wirkung. Dabei ist anzunehmen, dass die Fertigkeit, Bilder und Texte integriert zu verarbeiten, eine kulturspezifische Kompetenz darstellt, die erlernt werden muss. So erwerben Kinder in der Grund-

Diese Veröffentlichung wurde ermöglicht durch Sachbeihilfen der Deutschen Forschungsgemeinschaft (Kennz.: SCHN 665/3-2 und BA 1461/7-1) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

schule die Fähigkeit zu lesen, aber auch viele visuelle Informationen können nur verstanden werden, wenn ein entsprechendes Vorwissen hinsichtlich der Verarbeitung von Bildern bereits vorhanden ist (Weidenmann, 1994). Dies gilt insbesondere für die sogenannten logischen Bilder wie Kreis-, Balken-, und Liniendiagramme oder thematische Karten, deren Darstellungskonventionen den Erwerb spezifischer Kompetenzen notwendig machen (Schnotz, 1994b). Darüber hinaus stellt auch die Verarbeitung realistischer Bilder nicht selten eine Herausforderung dar, selbst wenn die abgebildeten Sachverhalte anhand von Schemata der alltäglichen Wahrnehmung erkannt werden (Weidenmann, 2006). Oft unterschätzen Lernende die Informationstiefe anschaulicher Zeichnungen oder Fotografien, was dazu führen kann, dass bildliche Repräsentationsformate nur oberflächlich verarbeitet werden. Aus diesem Grund enthalten Bilder in der Regel Beschriftungen oder sind in einen Text eingebettet, der das Verständnis der visuellen Informationen ermöglichen bzw. erleichtern soll. Ein Lernender muss jedoch in der Lage sein, die Inhalte des Bildes und des Textes aufeinander zu beziehen und integriert zu repräsentieren (Schnotz & Horz, 2009). Allerdings wird die integrative Verarbeitung von Texten und Bildern trotz ihrer Bedeutung für die alltäglichen Anforderungen in einer multimedialen Wissensgesellschaft an Schulen bislang kaum systematisch vermittelt (McElvany et al., 2009; Schroeder et al., in press). Es erstaunt daher wenig, dass in Untersuchungen zum Verstehen von Diagrammen in der Sekundarstufe I die Schülerinnen und Schüler teilweise gravierende Schwächen zeigten (Gobbo, 1994). Um Abhilfe zu schaffen, sind Konzepte und konkrete Maßnahmen erforderlich, die Lehrer und Auszubildende in die Lage versetzen, ihre Schüler bei der integrativen Verarbeitung von Text- und Bildinformationen anzuleiten und gezielt zu fördern. Allerdings ist bisher kaum etwas darüber bekannt, wann, wie und unter welchen Umständen sich diese Schülerkompetenz entwickelt. Eine Antwort auf diese Fragen soll das Projekt BiTe geben, das durch seine thematische Ausrichtung diese Lücke theoretisch und praktisch zu schließen versucht (vgl. Schnotz et al., 2010). Im Zentrum steht dabei die Entwicklung und Überprüfung eines Kompetenzmodells zur Verarbeitung von Text- und Bildinformationen, das neben Strukturen und Teilprozessen Aspekte der Kompetenzentwicklung umfasst¹.

Theoretische Ableitung eines Kompetenzmodells zur Text-Bild-Integration

Es existieren verschiedene theoretische Konstrukte, die den kognitiven Prozessen beim Lernen mit Texten und Bildern gewidmet sind (Mayer, 2005; Paivio, 1986). Viele An-

¹ Das ist jedoch nur die eine Hälfte des Projekts. Parallel wird untersucht, über welche Kenntnisse die Lehrkräfte im Umgang mit Texten und Bildern verfügen und inwieweit ihr Handeln im Unterricht von Ausbildung und Berufserfahrung abhängig ist. Die folgenden Ausführungen beziehen sich dagegen ausschließlich auf das Kompetenzmodell der Schülerebene.

sätze gehen dabei von einer parallelen Verarbeitung der verbalen und bildlichen Informationen aus, obwohl die beiden Formate auf unterschiedlichen Repräsentationsprinzipien beruhen. Während die Bedeutung von Texten durch die arbiträre Zuordnung von Symbolen bzw. Symbolketten zu Inhalten entsteht, besitzen Bilder in der Regel einen analogen Charakter, wodurch sich die dargestellten Sachverhalte unmittelbar ablesen oder erkennen lassen. Im Modell des integrativen Text- und Bildverstehens von Schnotz und Bannert (1999, 2003) wird eine solche Unterscheidung zwischen *deskriptionalen* und *deiktionalen* Repräsentationen vorgenommen, was konsequenter Weise formatspezifische Verarbeitungsprozesse nach sich zieht (siehe Abbildung 1). Demnach wird beim Verstehen eines Textes durch verbale Organisationsprozesse zunächst eine mentale Textoberflächenrepräsentation aufgebaut. Auf dieser Ebene werden syntaktische und morphologische Charakteristika des Textes mental abgebildet, aber noch keine Bedeutungen zugewiesen oder generiert. Eine Repräsentation des semantischen Gehalts entsteht erst durch konzeptuelle Organisationsprozesse, welche die Textoberfläche zu einem propositionalen Gefüge weiterverarbeiten. Dieses propositionale Gefüge bildet dann die Textbasis, auf deren Grundlage es möglich ist, ein mentales Modell der beschriebenen Sachverhalte zu konstruieren (Schnotz, 1994a; van Dijk & Kintsch, 1983). Sämtliche Konstruktionsprozesse erfolgen dabei vor dem Hintergrund auf- und absteigender Schemata, die als Vorwissen bei Bedarf aus dem Langzeitgedächtnis abgerufen werden (Anderson & Pearson, 1984; Schnotz, 1994a).

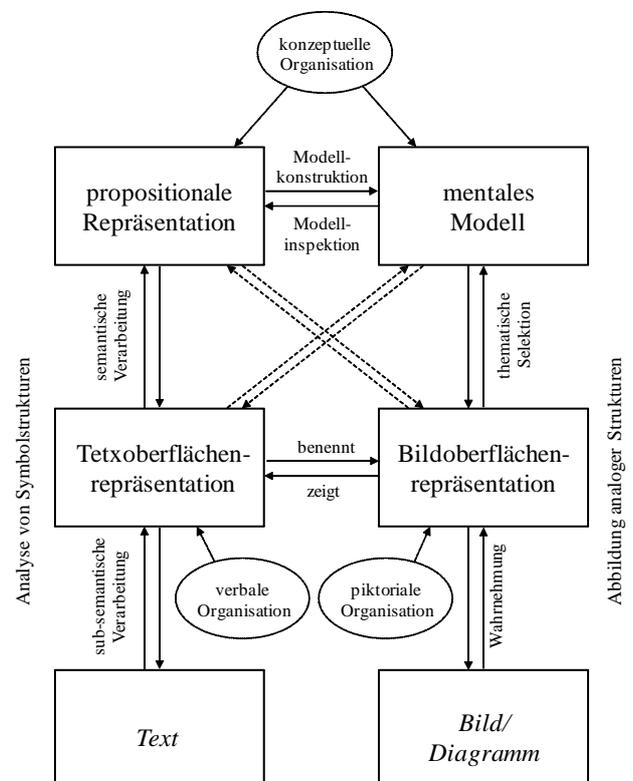


Abbildung 1. Modell des integrierten Text- und Bildverstehens (nach Schnotz & Bannert, 1999).

Beispielsweise können übergeordnete bzw. hierarchie-höhere Schemata im Sinne von Kontextinformationen genutzt werden, um ein unleserliches Wort zu identifizieren. Gleichzeitig aktivieren oder hemmen einzelne Worte übergeordnete Schemata, so dass sich letztlich die Konfiguration kognitiver Schemata herausbildet, die am besten zu den Textinformationen passt.

Ähnlich verhält es sich mit der Verarbeitung von visuellen Inhalten. So selektiert der Betrachter eines Bildes zunächst anhand absteigender Schemaaktivierungen einzelne Informationen, die durch automatisierte visuelle Routinen enkodiert werden (Ullman, 1984). Beim Anblick eines logischen Bildes würde zum Beispiel das allgemeine Schema Liniendiagramm aktiviert werden, das die Verarbeitung steuert. Zentrale Bestandteile des Enkodierungsvorgangs sind dabei Prozesse, bei denen einzelne grafische Elemente diskriminiert und identifiziert werden, um sie anschließend in übergeordneten Strukturen zu organisieren (Wertheimer, 1923; Winn, 1994). Auf diese Weise entsteht eine mentale Repräsentation, die Schnotz und Bannert (1999) als visuelle Wahrnehmung bzw. als Vorstellung bezeichnen. Auf der Ebene der visuellen Wahrnehmung werden vor allem perzeptive Strukturen, aber noch keine wirklichen Bedeutungen repräsentiert. Echtes Verstehen erfordert semantische Verarbeitungsprozesse, welche die wahrnehmungsnahe Repräsentation in ein mentales Modell überführen. Dies geschieht, indem Elemente und räumliche Relationen der visuellen Wahrnehmung auf mentale Elemente und semantische Relationen abgebildet werden. Nach Ansicht von Schnotz und Bannert (1999) ist Bildverstehen damit ein Prozess der schemageleiteten analogen Strukturabbildung, bei dem ein System visuell-räumlicher Relationen auf ein System von semantischen Relationen übertragen wird. Das entstehende mentale Modell besitzt deshalb nur noch strukturelle oder funktionale Ähnlichkeiten mit den Sachverhalten, die es repräsentiert. Zudem wird während der Konstruktion das mentale Modell anhand von Schemata aus dem Langzeitgedächtnis elaboriert, so dass es letztlich mehr Informationen umfassen kann, als die ursprüngliche Wahrnehmung.

Generell ist es ein Kennzeichen des Modells des integrierten Text-Bild-Verstehens, dass jeder Repräsentationsform depiktionale oder deskriptionale Informationen zugrunde liegen können. Der Betrachter eines Bildes wird in der Regel keine Probleme haben, seine visuellen Sinnesindrücke verbal zu artikulieren, während der Leser sich anschauliche Sachverhalte in einem Text bildlich vorstellen kann. Diese theoretischen Überlegungen haben zur Folge, dass Lernende beim gemeinsamen Wissenserwerb mit Texten und Bildern zwei Informationsquellen komplementär nutzen können. Da es jedoch möglich ist, das mentale Modell sowohl anhand der verbalen als auch anhand der piktoralen Informationen zu konstruieren, müssen sie sich entscheiden, auf welches Format, wann zurückgegriffen wird. Es ist davon auszugehen, dass bei dieser Entscheidung der kognitive Aufwand eine wichtige Rolle spielt. Dieser ist vergleichsweise gering, wenn die Konstruktion des mentalen Modells anhand eines Bildes stattfindet, da hier ein analoges Format in ein anderes analoges Format

überführt wird. Die Anforderungen, mit denen der Aufbau eines mentalen Modells auf der Grundlage eines Textes verbunden ist, sind dagegen deutlich höher. Das liegt vor allem an der Umwandlung einer propositionalen Struktur, die auf symbolischen Repräsentationsprinzipien beruht, in das analoge Format eines mentalen Modells.

Die eigentliche Integration der verbalen und bildlichen Informationen findet damit vor allem auf der semantischen Ebene statt und lässt sich als Interaktion zwischen der propositionalen Repräsentation und dem mentalem Modell beschreiben. Hier werden Strukturen der einen Repräsentationsform auf die Strukturen der anderen Repräsentationsform abgebildet oder *gemappt*, wobei sich verschiedene Strukturabbildungsprozesse unterscheiden lassen. Die einfachste Form beinhaltet das *mapping* von einzelnen Datenpunkten, bei dem eine referenzielle Verknüpfung zwischen einer Proposition und der Entität eines mentalen Modells hergestellt wird. Das ist beispielsweise dann der Fall, wenn ein unbekanntes Element in einer thematischen Karte anhand der Legende identifiziert werden muss. Beinhaltet der Strukturabbildungsprozess mehrere Elemente, die in einer semantischen Beziehung zu einander stehen, dann ist es erforderlich, eine referenzielle Verknüpfung zwischen Relationen herzustellen. Je nach Anzahl der Elemente und der Komplexität ihrer strukturellen Beziehungen können verschiedene Anforderungsniveaus unterschieden werden (vgl. Hegarty, Meyer & Narayanan, 2002). Das Mapping einfacher Relationen erfordert die Repräsentation von zwei Datenpunkten und deren Beziehung. Ein solcher Verarbeitungsprozess ist beispielsweise notwendig, wenn zwei Säulen in einem Diagramm miteinander verglichen werden sollen. Indem die Relation zwischen den Merkmalsträgern (Säulen) identifiziert und auf die jeweils andere Repräsentation abgebildet wird, entsteht eine kohärente Repräsentation dieser Beziehung, an der sich beispielsweise ablesen lässt, dass der eine Merkmalsträger größer als der andere ist. Das Mapping komplexer Relationen umfasst dagegen nicht nur die Verknüpfung von drei oder mehr Elementen bzw. Entitäten und die Repräsentation von mindestens zwei Relationen. Bei diesem Prozess müssen Beziehungen zwischen Beziehungen identifiziert und auf die andere Repräsentation abgebildet werden. Diese Anforderung stellt sich beispielsweise dann, wenn Schülerinnen und Schüler den Verlauf zweier Kurven in einem Liniendiagramm inhaltlich beschreiben und interpretieren müssen.

Jeder dieser *Mapping*-Prozesse setzt voraus, dass alle involvierten Elemente und Relationen gleichzeitig im Arbeitsgedächtnis präsent sind. Deshalb ist davon auszugehen, dass Integrationsprozesse, die viele Elemente und Relationen umfassen, auf einer höheren Anforderungsebene anzusiedeln und demnach schwieriger sind, als die Integrationsprozesse mit wenigen Elementen und nur einer Relation. Diese Überlegungen führen dazu, dass sich die Kompetenz, Bilder und Texte integriert zu verarbeiten, in drei strukturell verschiedene Teilkompetenzen unterteilen lässt. Diese Teilkompetenzen (TK) besitzen im Rahmen von Lern- und Entwicklungsprozessen eine Ermöglichungsfunktion, indem sie die Ausbildung der jeweils nächst höheren Teilkompetenz erlauben. Demnach ist das

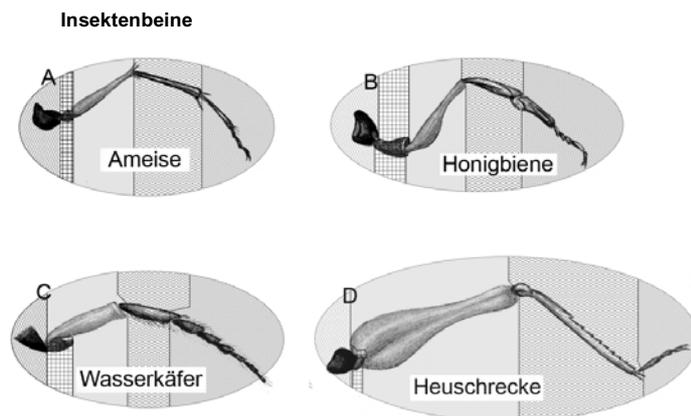
mappen einzelner Datenpunkte (TK1) die logische Voraussetzung für die Repräsentation einfacher Relationen (TK2) und das *mappen* einfacher Relationen ist die Voraussetzung für die Repräsentation komplexer Relationsgefüge (TK3). Aufgrund dieser strukturellen Beziehung sollten die drei Teilkompetenzen mittel bis hoch korreliert sein. Zudem ist mit dem Modell die Annahme verbunden, dass sich die Entwicklung der Integrationskompetenz als eine Sequenz von Verschiebungen auf drei abgrenzbaren Dimensionen darstellen lässt.

Theoriegeleitete Aufgaben- und Itemkonstruktion

Um die theoretischen Annahmen prüfen zu können, musste ein Instrument entwickelt werden, das den Spezifikationen des Kompetenzmodells entspricht. Als Vorlage dienten dabei Schulbücher der Fächer Biologie und Geografie der Klassenstufen fünf bis acht aus verschiedenen Bundesländern. Jede Aufgabe besteht aus einem Aufgabenstamm, der einen kurzen Text und ein bis drei Bildern (Karten, realistische Abbildungen sowie Kreis-, Balken- und Liniendiagramme) umfasst, sowie sechs Multiple-Choice-Fragen (Items) mit vier Antwortalternativen, von denen jeweils eine richtig ist. Alle Items wurden nach den Vorgaben des theoretischen Modells konstruiert, so dass pro Aufgabe immer zwei Items eine der drei Teilkompetenzen erfassen. Der Aufgabenstamm ist so gestaltet worden, dass Text- und Bildinformationen aufeinander bezogen werden müssen, wenn man die Fragen richtig beantworten will. Abbildung 2 zeigt das Beispiel eines Aufgabenstamms, der einem Biologiebuch entnommen wurde.

Wie man unschwer erkennen kann, geht es hier um die Beine von vier Insekten, die unterschiedliche Funktionen

besitzen. Eines der beiden Items mit der ersten kognitiven Teilkompetenz lautet daher: „Welches Insekt hat ein Putzbein?“ Zur Beantwortung dieser Frage muss ein Element aus dem Text (Putzbein, D) auf die Kennzeichnung des Bildes übertragen werden. Erst dann lässt sich das Tier bestimmen, zu dem das entsprechende Bein gehört. Bei einem Item, das zur Erfassung der zweiten Teilkompetenz entwickelt worden war, wird gefragt: „Welche Art von Bein hat das Insekt mit der längsten Schiene?“ Die Lösung erfordert zunächst die Identifikation des entsprechenden Körperteils, das im Original durch ein grelles Rosa gekennzeichnet ist. Diese Bedeutung kann wiederum dem Text entnommen und auf das Bild übertragen werden. Allerdings wird nach der längsten Schiene gefragt, was die mentale Repräsentation von einfachen Relationen verlangt, die solange paarweise gebildet werden, bis das Sprungbein der Heuschrecke als richtige Antwort identifiziert ist. Etwas komplizierter wird es bei der dritten Anforderungsebene, wenn der Satz „Das Putzbein hat im Vergleich zum Sprungbein ...“ vervollständigt werden muss. Um die richtige Lösung („... einen dünneren Schenkel aber einen längeren Fuß“) zu finden, die sich wie die übrigen Antwortalternativen, auf jeweils zwei unterschiedliche Beinsegmente beziehen, ist eine ganze Reihe von Prozessen erforderlich. Zuerst müssen anhand der Farbkodierung im Text die betreffenden Segmente identifiziert werden. Grundlage für die Prüfung der einzelnen Aussagen ist dann die eigene mentale Repräsentation des Sachverhaltes. Die Anforderungen dafür sind relativ hoch, da gleichzeitig vier Beinsegmente und deren Relationen im Arbeitsgedächtnis repräsentiert werden müssen. Erschwerend kommt hinzu, dass es bei der Bearbeitung der Items nicht erlaubt ist, sich Notizen zu machen, um die kognitive Belastung zu reduzieren. Bei dem angeführten Beispiel handelt es sich um eine vergleichsweise leichte Aufgabe, deren Items von einem Großteil der Schüler einer fünften Klasse erfolgreich gelöst werden konnte.



Die Beine der Insekten in den Abbildungen A - D haben alle den gleichen Grundaufbau:

Hüfte , Schenkelring , Schenkel , Schiene und Fuß .

Die Beine sind vor allem Fortbewegungsorgane, die zum Laufen (Laufbein, A), Schwimmen (Schwimmbein, B) oder Springen (Sprungbein, C) dienen. Sie können aber auch zum Putzen (Putzbein, D) verwendet werden.

Abbildung 2. Beispiel für einen Aufgabenstamm.

Der Untersuchungsansatz

Ein wichtiger Bestandteil des Modells sind Annahmen zum Entwicklungsverlauf der Kompetenz Bilder und Texte integrativ zu verarbeiten. Aus diesem Grund wurde ein längsschnittlicher Versuchsplan umgesetzt, bei dem zwei Schüler-Kohorten zu je 24 Klassen über einen Zeitraum von drei Jahren begleitet werden. Während in der Grundschule die Unterrichtsmaterialien vor allem Bilder mit situierender oder dekorativer Funktion enthalten, verändern sich die Anforderungen mit dem Übergang in die Sekundarstufe. Hier werden in vielen Fächern (z. B. Biologie oder Erdkunde) Bilder verwendet, die eine instruktionale Funktion besitzen und als eigenständige Informationsquelle zusammen mit dem Text zu verarbeiten sind. Da der Sekundarstufe somit inhaltlich eine große Bedeutung zukam, sollte die Längsschnittsuntersuchung in den Klassen fünf und sechs beginnen.

Für die Umsetzung des Versuchsplans wurden drei bis vier Testhefte benötigt, die eine Wiederho-

lungsmessung ermöglichen sollten, ohne den Schülern die gleichen Aufgaben zweimal vorzulegen. Daher wurde zunächst eine Pilotierungsstudie durchgeführt, bei der eine Stichprobe von 48 Klassen ($N = 1060$) der Stufen fünf bis acht insgesamt 48 Aufgaben bearbeitete. Die Grundgesamtheit, aus denen die teilnehmenden Klassen zu gleichen Teilen zufällig gezogen wurden, umfasste sämtliche Hauptschulen, Realschulen und Gymnasien des Landes Rheinlandpfalz. Da aber ein einzelnes Kind nicht in der Lage ist, 48 Aufgaben auf einmal zu bearbeiten, erhielt jeder Schüler ein Testheft mit einer Auswahl von insgesamt 8 Aufgaben. Um dennoch alle Items auf einer gemeinsamen Skala anordnen zu können, wurden die Testhefte nach einem Youden-Square-Design zusammengestellt, mit dem eine Verlinkung der Aufgaben verwirklicht werden konnte (Hambleton, Swaminathan & Rogers, 1991). Ein solcher Ansatz erforderte jedoch die Anwendung eines Verfahrens der probabilistischen Testtheorie, weshalb die Skalierung anhand des Raschmodells erfolgte. Zu dessen Grundannahme zählt, dass die konkreten Antworten in einem Test von der Schwierigkeit der Fragen bzw. Aufgaben (Itemparameter) und der Ausprägung eines latenten Persönlichkeitsmerkmals (Personenparameter) abhängen. Beide Parameter lassen sich allerdings nicht direkt beobachten, weshalb diese auf der Basis der Antwortmuster durch spezielle Varianten der Maximum-Likelihood-Methode geschätzt werden müssen (vgl. Rost, 2004). Im Rahmen der BiTe-Studie wurde für die Parameterschätzung das Computerprogramm ConQuest (Wu, Adams, Wilson & Haldane, 2007) verwendet, das unter anderem die Möglichkeit bietet mehrdimensionale Raschmodelle zu skalieren.

Prüfung der Kompetenzstruktur

Um zu klären, wie viele Dimensionen dem Konstrukt zugrunde liegen, wurden die Items anhand von alternativen Testmodellen kalibriert und die Passung miteinander verglichen. Damit sollte vor allem überprüft werden, ob die drei Teilkompetenzen jeweils eigenständige Dimensionen konstituieren, oder ob alle Items auf einer Dimension zu verorten sind. Da die Repräsentation und Strukturabbildung von Relationen eine besondere Herausforderung darstellt, wurde zusätzlich ein Testmodell untersucht, bei dem die Items der Teilkompetenz eins die erste Dimension und die Items der Teilkompetenzen zwei und drei gemeinsam eine zweite Dimension bilden. Zur Einschätzung, welches der Modelle die Antwortmuster der Schüler am besten beschreibt, wurde das Bayes-Information-Kriterium (BIC) bestimmt. Tabelle 1 gibt die Ergebnisse für die verschiedenen a priori definierten Testmodelle wieder. Dabei

ist zu erkennen, dass die zweidimensionale Lösung den niedrigsten Wert aufwies und offensichtlich am besten auf die Daten passte, während für das dreidimensionale Modell der schlechteste BIC-Wert ermittelt wurde.

Berücksichtigt man aber die Korrelation zwischen den beiden Skalen des zweidimensionalen Modells, die mit $r = .76$ sehr hoch ist und messfehlerbereinigt bei $r = .95$ liegt, kann anhand des Ergebnisses nicht ausgeschlossen werden, dass es sich doch um ein eindimensionales Konstrukt handelt. Ein Modellvergleich anhand von informationstheoretischen Maßen ist keine statistische Absicherung, was eine Interpretation des Ergebnisses zusätzlich erschwert. Somit lässt sich anhand der Resultate in Tabelle 1 nicht feststellen, wie groß die Passungsunterschiede zwischen den drei Modellen sind. Eine Alternative stellt der Likelihoodquotiententest dar, doch bei umfangreichen Instrumenten mit vielen Items wird von einer Anwendung abgeraten, da die asymptotische Bedingung der χ^2 -Prüfstatistik als verletzt gilt (Rost, 2004). Um dennoch eine Grundlage für die Entscheidung für oder wider eines der Modelle zu schaffen, wurde die zweidimensionale Lösung ein weiteres Mal geschätzt. Bei dieser Schätzung war jedoch die latente Korrelation zwischen den beiden Skalen zuvor auf den Wert $r = 1.0$ fixiert worden. Der resultierende BIC-Wert liegt mit 46821.8 unter dem Wert für die zweidimensionale Schätzung ohne fixierte Korrelation, weshalb das Modell mit der perfekten latenten Korrelation am besten auf die Daten passt. Das Ergebnis wurde so interpretiert, dass beide Skalen die gleiche latente Variable erfassen. In diesem Zusammenhang stellt sich allerdings die Frage nach den Gründen für den unterschiedlichen Modellfit der eindimensionalen und der restringierten zweidimensionalen Lösung. Die Abweichung lässt sich damit erklären, dass aufgrund der Verwendung von ConQuest (Wu et al., 2007) die Schätzung der Itemparameter auf der Marginal Maximum Likelihood-Methode beruht. Bei diesem Verfahren werden neben der Schwierigkeit der Items unter anderem auch die Mittelwerte der latenten Variablen geschätzt. Den Schülerinnen und Schülern fiel es jedoch leichter, einzelne Datenpunkte zu identifizieren, als Relationen zu bilden, weshalb die zweidimensionale Lösung trotz Restriktion zwei Mittelwerte ergab, die deutlich voneinander abweichen. Die Items der beiden Dimensionen erfassen zwar das gleiche Konstrukt, allerdings auf verschiedenen Anforderungsebenen. Dieser Unterschied kann durch das eindimensionale Modell nicht abgebildet werden, da dessen Schätzung nur eine latente Variable und damit nur einen Mittelwert umfasst. Die gute Passung der zweidimensionalen Struktur ist damit vor allem auf die Schwierigkeitsunterschiede zwischen den Items der Teilkompetenzen zurückzuführen.

Tabelle 1. Ergebnisse der Modellanpassungsprüfung für Testmodelle mit ein, zwei und drei Dimensionen

Testmodell	-2 ln L	Parameter	BIC
1-dimensional	44841.7	289	46854.9
2-dimensional	44819.3	291	46846.4
3-dimensional	44815.6	294	46863.7

Diskussion und Ausblick

Aufgrund dieser Ergebnisse wurden die Items der Pilotierungsuntersuchung schließlich anhand des eindimensionalen Raschmodells kalibriert. Dabei zeigte sich, dass etwa zwei Drittel der Aufgaben als modellkonform eingestuft werden konnten, so dass es möglich war, vier klassenspezifische Testhefte gleicher Länge zusammenzustellen. Mittlerweile ist auch die Hauptuntersuchung angelaufen, doch ist die Frage nach der Struktur der Kompetenz zur integrativen Text-Bild-Verarbeitung noch nicht endgültig geklärt. Festzuhalten bleibt, dass die Skalen der mehrdimensionalen Modelle hoch korreliert sind, was eine Abgrenzung der Teilkompetenzen und damit auch der alternativen Testmodelle erschwert haben dürfte. Dieses Ergebnis konnte jedoch nicht überraschen, da nach der theoretischen Herleitung des Kompetenzmodells mit einem solchen Zusammenhang gerechnet werden musste. Um ein Item der dritten Teilkompetenz zu lösen, ist es demnach nicht nur erforderlich, Relationen zweiter Ordnung herstellen zu können. Konstituierend sind die hierarchieniederen Anforderungen, auf denen der komplexere Prozess aufbaut. Beispielsweise müssen Schülerinnen und Schüler zunächst in der Lage sein, Datenpunkte durch Integration der Text- und Bildinformationen abzulesen, bevor sie eine Relation mental repräsentieren können. Dies ist wiederum die Voraussetzung für die mentale Repräsentation von Relationen zwischen Relationen. Ein Item der dritten Teilkompetenz umfasst demnach zum Teil die gleichen Anforderungen, wie Items der anderen beiden Teilkompetenzen, so dass eine hohe Korrelation auch theoretisch zwingend ist.

Dabei stellt sich die Frage, ob Testmodelle, die von eindeutig separierbaren Dimensionen ausgehen, der Struktur der Daten gerecht werden können. Die Antwort lautet mit hoher Wahrscheinlichkeit nein. So lässt sich bei der Schätzung mehrdimensionaler IRT-Modelle eine *between-item-dimensionality* von einer *within-item-dimensionality* unterscheiden (Wu et al., 2007). Im ersten Fall gehören die Items genau einer Dimension an, während bei Modellen mit *within-item-dimensionality* sich die Items auf mehr als eine Dimension beziehen. Nach dem theoretischen Modell trifft genau das auf die drei postulierten Teilkompetenzen zu, deren Prozesse ineinander verschachtelt sind. Die Schätzung von Modellen mit Items, die auf mehreren latenten Dimensionen gleichzeitig laden können, ist jedoch alles andere als trivial. Sind in einem Datensatz zu viele systematische Missing-Werte oder ist die Anzahl der Items zu hoch, kommt es relativ schnell zu Problemen mit den Schätzalgorithmen, die äußerst rechenintensiv sind. Aus diesem Grund war es bislang nicht möglich, Modelle mit *within-item-dimensionality* auf die Daten der Pilotierungsstudie anzuwenden.

Eine Alternative stellt das linear-logistische Testmodell nach Fischer (1983) dar, bei dem anstelle von Items kognitive Teilprozesse definiert und geschätzt werden. Die Möglichkeiten dieses Modells waren Anlass für eine kognitive Itemanalyse, die zurzeit ausgewertet wird. Unter anderem war es das Ziel, die Anforderungen der einzelnen Teilkompetenzen noch weiter zu spezifizieren, damit zu-

sätzliche Charakteristika und Prozesse, die in der ursprünglichen Definition nicht enthalten sind, berücksichtigt werden konnten. Das betraf unter anderem Wechselwirkungen der Items mit dem jeweiligen Aufgabenstamm, was beispielsweise die Bildart oder die Textlänge miteinschloss. Eine solch feinkörnige Strukturanalyse kann der Ausgangspunkt für die Entwicklung von Interventionsstrategien zur Förderung der integrativen Text-Bild-Verarbeitung sein. Dies gilt umso mehr, wenn die Ergebnisse der Schüler- und der Lehrerebene des BiTe-Projekts zusammgeführt werden.

Literatur

- Anderson, R. C. & Pearson, P. D. (1984). A schema-thematic view of basic processes in reading comprehension. In P. D. Pearson, R. Barr, M. L. Kamil & P. Mosenthal, (Eds.), *Handbook of reading research* (pp. 255–291). New York, NY: Longman.
- Carney, R. N. & Levin, J. R. (2002). Pictorial illustrations still improves students' learning from text. *Educational Psychology Review*, 14 (1), 5–26.
- Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, 48, 3–26.
- Gobbo, C. (1994). On children's understanding of an economic concept: the role of graphics in evaluation. In W. Schnotz & R. Kulhavy (Eds.), *Comprehension of graphics* (pp. 227–249). Amsterdam: North Holland.
- Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hegarty, M., Meyer, B. & Narayanan, N. H. (Eds.). (2002). *Diagrammatic representation and inference*. Berlin: Springer.
- Levie, H. W. & Lentz, R. (1982). Effects of text illustration: A review of research. *Educational Communication and Technology Journal*, 30, 195–232.
- Levin, J. R., Anglin, G. J. & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (pp. 51–85). New York, NY: Springer.
- Mayer, R. E. (Ed.). (2005). *The cambridge handbook of multimedia learning*. Cambridge, UK: Cambridge University Press.
- McElvany, N., Schroeder, S., Richter, T., Hachfeld, A., Baumert, J., Schnotz, W. et al (2009). Diagnostische Fähigkeiten von Lehrkräften bei der Einschätzung von Schülerfähigkeiten und Aufgabenschwierigkeiten bei Lernmedien mit instruktionalen Bildern. *Zeitschrift für Pädagogische Psychologie*, 23, 223–235.
- Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York, NY: Oxford University Press.
- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion* (2., vollständig überarbeitete und erweiterte Aufl.). Bern: Huber.
- Schnotz, W. (1994a). *Aufbau von Wissensstrukturen. Untersuchungen zur Kohärenzbildung bei Wissenserwerb mit Texten*. Weinheim: Beltz PVU.
- Schnotz, W. (1994b). Wissenserwerb mit logischen Bildern. In B. Weidenmann (Hrsg.), *Wissenserwerb mit Bildern* (S. 95–147). Bern: Huber.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 49–69). Cambridge, UK: Cambridge University Press.
- Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Zeitschrift für Experimentelle Psychologie*, 46 (3), 217–236.

- Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13*, 141–156.
- Schnotz, W. & Horz, H. (2009). Online-Lernen mit Texten und Bildern. In L. J. Issing & P. Klimsa (Hrsg.), *Online-Lernen* (S. 87–104). München: Oldenbourg.
- Schnotz, W., Baumert, J., Horz, H., McElvany, N., Ullrich, M., Schroeder, S. et al. (2010). Das BITE-Projekt: Integrative Verarbeitung von Texten und Bildern in der Sekundarstufe I. *Zeitschrift für Pädagogik [Beiheft] 56*, 143–153.
- Schroeder, S., Richter, T., McElvany, N., Hachfeld, A., Baumert, J., Schnotz, W. et al. (in press). Effects of teachers' beliefs and instructional behaviors on students' engagement in learning from texts with pictures. *Learning and Instruction*.
- Ullman, S. (1984). Visual routines. *Cognition, 18*, 97–159.
- van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York, NY: Academic Press.
- Weidenmann, B. (1994). *Wissenserwerb mit Bildern*. Bern: Huber.
- Weidenmann, B. (2006). Lernen mit Medien. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (5. Aufl., S. 423–476). Weinheim: Beltz.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt, II. *Psychologische Forschung, 4*, 301–350.
- Winn, W. D. (1994). Contributions of perceptual and cognitive processes to the comprehension of graphics. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of graphics* (pp. 3–27). Amsterdam: North Holland.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest Version 2.0. Generalised Item Response Modelling Software*. Camberwell, VIC: Acer Press.

Dipl.-Psych. Mark Ullrich

Goethe Universität Frankfurt
Mertonstraße 17
60325, Frankfurt
E-Mail: m.ullrich@psych.uni-frankfurt.de