

Originalia

Entwicklung und erste Validierung einer Skala zur Erfassung der passiven argumentativ-rhetorischen Kompetenz

Development and First Validation of a Scale of Passive Argumentational-Rhetorical Competence (SPARK)

Jürgen Flender

Universität Heidelberg

Ursula Christmann

Universität Heidelberg

Norbert Groeben

Universität Köln

Zusammenfassung: Unter passiver argumentativ-rhetorischer Kompetenz verstehen wir die Fähigkeit zur Identifikation von argumentativen, rhetorischen, argumentationslogischen und interaktiven Auffälligkeiten in einer Argumentation. Die «Skala zur Erfassung der passiven argumentativ-rhetorischen Kompetenz» (SPARK) erfaßt diese vier Kompetenzaspekte mit Hilfe eines Szenario-Ansatzes. Nach dem Kriterium der inhaltlichen Validität wurden zwei parallele Formen mit je fünf Szenarios entwickelt. Ergebnisse einer Validierungsstudie mit 116 Untersuchungspartnern/innen (Uptn) zeigen eine Korrelation von .75 zwischen den beiden inhaltlich parallelen Versionen, geringe und mäßige Korrelationen zwischen den vier Konstruktdimensionen sowie geringe Korrelationen zwischen den Szenarios. Cronbachs α beträgt für Version A .68, für Version B .71. Uneinheitliche Werte werden in Übereinstimmung mit der inhaltlichen Vielschichtigkeit des Konstrukts, der inhaltsvaliden Konstruktion von SPARK sowie den deutlichen, für einen Niveautest aber noch angemessenen Schwierigkeitsunterschieden zwischen den Szenarios interpretiert.

Schlüsselwörter: Argumentation, Rhetorik, Argumentationsintegrität, argumentativ-rhetorische Kompetenz

Abstract: This article describes the development and validation of a scale for the assessment of passive argumentational-rhetorical competence (SPARK) which has been defined as the ability to identify conspicuous argumentational, rhetorical, logical and interactional aspects in contributions to argumentative discussions. According to the criterion of content validity SPARK has been developed in two parallel forms; each form consists of five (written) argumentational episodes which have to be analyzed with regard to the conspicuous aspects differentiated above. The results of a study with 116 subjects show a correlation of .75 between the two versions, weak and moderate correlations between the four dimensions of the construct and weak correlations between the argumentational episodes. The values for Cronbachs α are .68 in form A and .71 in form B. Inconsistent results are explained by the complexity of the construct and its heterogeneity on the content level, the content valid test construction and the clear, but acceptable differences in difficulty between the argumentational episodes.

Keywords: Argumentation, rhetorics, argumentational integrity, argumentational-rhetorical competence

Das Konzept der passiven argumentativ-rhetorischen Kompetenz ist in der Literatur bislang nicht als eigenständiges Konstrukt ausgearbeitet worden. Eine Reihe von Arbeiten befaßt sich mit kognitiven ([Kuhn, 1991](#)) bzw. allgemeinen entwicklungspsychologischen Voraussetzungen und Bedingungen von moralischem ([Miller, 1980, 1982, 1984](#)) bzw. kindlichem Argumentieren ([Koehler, 1992](#); [Overton, 1990](#); [Stein & Miller, 1989](#); [Völzing, 1982](#); zu einer Übersicht über Stufenmodelle der Entwicklung argumentativen Denkens und Verhaltens vgl. auch [Hofer, Fleischmann & Pikowsky, 1991](#)). Zu den herausgearbeiteten Voraussetzungen argumentativer Kompetenz gehört dabei unter entwicklungspsychologischer Perspektive insbesondere die Überwindung von egozentrischem Denken (sensu Piaget) sowie die Fähigkeit zur sozialen Perspektivenübernahme. Diese Voraussetzungen sind etwa ab dem elften Lebensjahr gegeben. Sie können daher für Erwachsene, auf die sich die Skala zur Erfassung passiver argumentativ-rhetorischer Kompetenz beziehen soll, als gegeben unterstellt werden. Für das Argumentieren Erwachsener liegen bislang keine entsprechenden theoretischen Ausarbeitungen vor.

Darüber hinaus mangelt es an geeigneten empirischen Instrumenten zur Erfassung passiver argumentativ-rhetorischer Kompetenz. Bei den bisher vorliegenden Erhebungsinstrumenten handelt es sich mehrheitlich um Fragebögen, die sich auf Neigungen der Sprecher/innen, beispielsweise auf deren Tendenz zum (aktiven) Argumentieren beziehen ([Infante & Rancer, 1982](#); vgl. auch [Rancer, Kosberg & Bauskus, 1992](#)). Insgesamt sind Kompetenzaspekte der Rezipienten/innen, wie beispielsweise deren Fähigkeit zur Diagnose unfairer Argumente, bislang weder theoretisch ausgearbeitet noch in entsprechende Instrumente umgesetzt worden. Auf eine Förderung solcher Kompetenzaspekte zielt ein

Trainingsprogramm des (von der DFG geförderten) Forschungsprojekts «Argumentationsintegrität» ab, das für unintegres Argumentieren sensibilisieren und eine Palette von integren Reaktionsmöglichkeiten auf unfaire Argumentationsbeiträge aufbauen will ([Groeben, Christmann & Mischo, 1997](#); [Mischo et al., 1999](#)). Für die Evaluation eines solchen Trainings, aber auch für die weitere Grundlagenforschung zur Qualität von Argumentationsstrukturen und -prozessen ist die Skala zur Erfassung der passiven argumentativ-rhetorischen Kompetenz entwickelt worden, die wir im folgenden vorstellen und abkürzend als «SPARK» bezeichnen.

Konstruktexplikation

Das Konstrukt der passiven argumentativ-rhetorischen Kompetenz wird gefaßt als die Fähigkeit zur Identifikation von argumentativen, rhetorischen und interaktiven Auffälligkeiten sowie als Fähigkeit, die Logik einer Argumentation nachvollziehen zu können. Entsprechend werden die folgenden vier Kompetenzbereiche (Konstruktdimensionen) unterschieden: passive argumentative Kompetenz, passive rhetorische Kompetenz, passive argumentationslogische Kompetenz und passive interaktive Kompetenz. Da es sich um ein neuartiges Konstrukt handelt, werden die vier Aspekte nachfolgend im einzelnen vorgestellt.

1. Passive *argumentative* Kompetenz bezeichnet die Fähigkeit, argumentative Auffälligkeiten zu identifizieren. Darunter werden Verstöße gegen Regeln fairen Argumentierens verstanden. Kriterien zur Feststellung solcher Verstöße bietet das Konstrukt der «Argumentationsintegrität» (eine differenziertere Konstruktexplikation geben: [Groeben, Schreier & Christmann, 1993](#); [Schreier, Groeben & Christmann, 1995](#)), dessen empirisch validiertes System aus Bedingungen, Merkmalen, Standards und Strategien (un-)integren Argumentierens (vgl. z. B. [Schreier, Groeben & Blickle, 1995](#)) die Grundlage für die inhaltsvalide Auswahl der argumentativen Auffälligkeiten in SPARK bildete; die dabei zentralen Strukturmerkmale sind folgende:

Das Konstrukt der «Argumentationsintegrität» expliziert auf hohem Abstraktionsniveau vier grundlegende Bedingungen fairen Argumentierens: (I) formale Richtigkeit, (II) inhaltliche Richtigkeit/Aufrichtigkeit, (III) inhaltliche Gerechtigkeit und (IV) prozedurale Gerechtigkeit/Kommunikativität sowie vier komplementäre Klassen argumentativer Regelverletzungen: I. fehlerhafte Argumentationsbeiträge, II. unaufrichtige Argumentationsbeiträge, III. ungerechte Argumente, IV. ungerechte Interaktionen (ausführlich: [Groeben, Schreier & Christmann, 1993](#)). Diesen Merkmalen wurden auf der Basis von empirischen Experten- und Laien-Kategorisierungen ethisch problematischer Strategien der Rhetorik 11 Standards des integren Argumentierens zugeordnet ([Schreier, 1994](#); [Schreier & Groeben, 1996](#)). Grundlage für die Strategieneinschätzung war eine Auswahl von 86 aus insgesamt 207 ethisch bedenklichen Strategien aus 45 populärrhetorischen Ratgebern ([Schreier, 1992](#)). [Tabelle 1](#) zeigt neben Bedingungen,

Merkmale und Standards beispielhaft diejenigen acht Strategien des unintegren Argumentierens, die sich in bisherigen Validierungs-Studien ([Schreier & Groeben, 1990, 1996](#)) als besonders typisch für einen Merkmalsbereich erwiesen haben; sie sind daher als aufzufindende argumentative Auffälligkeiten in SPARK aufgenommen worden.

2. Passive *rhetorische* Kompetenz bezieht sich auf die Fähigkeit zur Identifikation von rhetorischen Auffälligkeiten. Damit sind Besonderheiten in der sprachlich-stilistischen Gestaltung gemeint, die von Argumentationsteilnehmern/innen zur Stützung ihrer Argumente eingesetzt werden. Dabei handelt es sich um rhetorische Stilfiguren wie beispielsweise Ironie, Verwendung von Metaphern oder parallele Satzstrukturen. Diese und viele weitere Figuren werden seit der Antike eingesetzt, um die Wirkung von integren Argumenten zu verändern. Inzwischen liegen empirische Befunde vor, die zeigen, daß eine gelungene ästhetische Gestaltung die persuasive Wirksamkeit von Argumenten erhöhen kann (vgl. [Mischo, Groeben & Christmann, 1997](#)). Der korrekten Identifikation solcher Stilmittel kommt daher eine hohe Bedeutung zu.
3. Passive *argumentationslogische* Kompetenz manifestiert sich im Verständnis für die Logik einer Argumentation. Ein solches Verständnis sollte sich in dem Ausmaß zeigen, in dem weitere Argumente im Hinblick auf die hypothetische Fortsetzung einer Argumentation in ihrer Stoßrichtung zutreffend eingeschätzt, d. h. der jeweils strittigen These als Pro-, Contra- oder aber als im vorliegenden Zusammenhang bedeutungsloses Argument zugeordnet werden.
4. Passive *interaktive* Kompetenz faßt die Identifikation von konfrontativen Äußerungen einerseits und gemeinsam geteilten Bezugspunkten andererseits zusammen. Als konfrontativ gelten Äußerungen, in denen gegensätzliche Standpunkte besonders deutlich, aber nicht notwendigerweise unfair zum Ausdruck kommen. Von Bedeutung ist das Identifizieren von Konfrontativität unter anderem dann, wenn es darum geht, das Ausmaß an Bewußtheit einzuschätzen, mit der ein argumentativer Regelverstoß begangen wird, denn Konfrontativität läßt sich als Indikator für Bewußtheit auffassen ([Christmann, Schreier & Groeben, 1996](#); [Schreier, 1997](#)). Einen weiteren Aspekt der Interaktion stellen gemeinsame Bezugspunkte dar. Dabei handelt es sich um Aussagen, die zwischen den Beteiligten unstrittig und deshalb aufgrund ihres impliziten Charakters häufig neu aus dem Gesprächszusammenhang zu erschließen sind. Die Fähigkeit, Strittiges und Unstrittiges zu unterscheiden, trägt zur Vermeidung von Mißverständnissen, unnötigen Standpunktklärungen etc. und damit zur Effizienz von Argumentationen bei; sie bildet somit einen bedeutsamen Ausschnitt passiver argumentativ-rhetorischer Kompetenz ab.

Die vier aufgeführten Aspekte sind als inhaltliche Kerndimensionen passiver argumentativ-rhetorischer Kompetenz definitorisch festgelegt worden. Unter dem Vorbehalt der empirischen Bewährung ist zusätzlich die Urteilssicherheit der Uptn als letzter Aspekt passiver argumentativ-rhetorischer Kompetenz einbezogen worden. Es wird erwartet, daß sich eine hohe passive argumentativ-rhetorische Kompetenz darin zeigt, daß zutreffende

Antworten mit hoher Urteilssicherheit, unzutreffende Antworten dagegen mit niedriger Urteilssicherheit gegeben werden. Bei der Vergabe von Punktzahlen wird Unsicherheit im Falle von unzutreffenden Antworten als Problembewußtsein interpretiert und mit höheren Punktzahlen bewertet als unzutreffende Antworten, die mit hoher Sicherheit gegeben werden. Das (vorläufige) Punktevergabemodell ist dabei so konstruiert, daß sich in keinem Fall falsche Antworten durch günstige Sicherheitsangaben kompensieren lassen. Darüber hinaus berücksichtigt das Punktevergabemodell die Vollständigkeit der Antwort: Maximale Punktzahlen erhalten diejenigen Uptn, die eine Auffälligkeit nicht nur korrekt lokalisieren, sondern diese anschließend auch (bei hoher Urteilssicherheit) zutreffend benennen. Zusammenfassend beinhaltet das Konstrukt der passiven argumentativ-rhetorischen Kompetenz die folgenden Bereiche:

1. Passive *argumentative* Kompetenz: Identifikation von argumentativen Regelverletzungen (Urteilssicherheit)
2. Passive *rhetorische* Kompetenz: Identifikation von rhetorischen Stilfiguren (Urteilssicherheit)
3. Passive *argumentationslogische* Kompetenz: Einschätzen weiterer Argumente (Urteilssicherheit)
4. Passive *interaktive* Kompetenz: Identifikation von konfrontativen Äußerungen und gemeinsam geteilten Aussagen (Urteilssicherheit)

Testkonstruktion

Im Hinblick auf die geplante Verwendung als Evaluationsinstrument ist SPARK unter dem vorrangigen Kriterium der inhaltlichen Validität als ein individueller Kompetenztest entwickelt worden, der auf unterschiedlichen Schwierigkeitsniveaus die Erfassung individueller Leistungsveränderungen erlaubt. Für die Verwendung als Forschungsinstrument sind darüber hinaus in einer Validierungsstudie Werte für die wichtigsten klassischen Testgütekriterien ermittelt worden (s. u.).

Methodisch folgt SPARK dem Szenario-Ansatz, der die Bewertung konstruierter Argumentationsepisoden (Szenarios) vorsieht. Im Falle von SPARK handelt es sich um verschriftete Szenarios, deren Zeilen durchnummeriert sind, um den Uptn die Angabe auffälliger Stellen zu ermöglichen. Insgesamt enthält SPARK die folgenden Aufgabentypen und Antwortmöglichkeiten: 1. Angabe der Zeilennummer zur Kennzeichnung auffälliger Stellen, 2. Multiple-Choice-Antworten zur Benennung von Auffälligkeiten, 3. freie Antworten zur fakultativen Angabe weiterer Auffälligkeiten und 4. Angaben zur Urteilssicherheit in fünffacher Abstufung.

SPARK wurde in zwei inhaltlich parallelen Formen (Versionen A und B) mit jeweils fünf

Szenarios entwickelt. Das Rahmenthema für Version A lautet: «Macht Fernsehen aggressiv?» Version B bezieht sich auf das Rahmenthema: «Ist Intelligenz angeboren?» Beide Rahmenthemen wurden von Experten/innen unter Themen ausgewählt, die von einer breiten Zielgruppe als vertraut, interessant und relevant empfunden werden sollten. Zusätzlich wurde eine Einführung in jeden der beiden Themenbereiche erstellt. Ziel dieser Maßnahme war es, eine gemeinsame Wissensbasis herzustellen und damit störenden Einflüssen aufgrund unterschiedlicher Vorkenntnisse, Interessen und Bildungsniveaus entgegenzuwirken. In die Szenarios wurden jeweils eine argumentative und zwei rhetorische Auffälligkeiten sowie eine konfrontative Äußerung eingearbeitet, wobei in jeder Version jeweils ein Szenario argumentativ unauffällig blieb. Zusätzlich wurden weitere potentielle Argumente sowie Aussagen generiert, die zwischen den beteiligten Sprecher/innen unstrittig sind. Alle Aufgaben wurden in einem differenzierten Punktevergabesystem integriert. Bezüglich der argumentativ-rhetorischen Auffälligkeiten stellte eine Ausbalancierung sicher, daß zwischen dem Geschlecht der Sprecher/innen und den von ihnen realisierten Auffälligkeiten kein systematischer Zusammenhang besteht. Zur Vermeidung von Positionseffekten wurden darüber hinaus die Auffälligkeiten sowohl innerhalb der Szenarios als auch innerhalb der vorgegebenen Antwortmöglichkeiten an unterschiedlichen Positionen plazierte. Die Prinzipien der Testkonstruktion sollen im folgenden am Beispiel von Szenario L aus Version A veranschaulicht werden. Szenario L enthält (1) Ironie und (2) Anapher als rhetorische Auffälligkeiten, (3) einen unzulässigen Umkehrschluß als argumentative Auffälligkeit sowie (4) Konfrontativität als interaktive Auffälligkeit; die jeweiligen Auffälligkeiten sind im nachfolgenden Beispielszenario unterstrichen. Im Original sind die Zeilen des Gesprächs numeriert.

Ein Gespräch im Lehrer/innen-Zimmer

Herr A und Frau B sitzen in einer Freistunde bei einer Tasse Kaffee zusammen. Beide unterrichten seit vielen Jahren gemeinsam an dieser Realschule im ländlichen Raum. Der gute Ruf ihrer Schule hat durch zunehmende Gewalt unter den Schülern/innen einige Kratzer bekommen. Soeben ist Frau B von der Pausenaufsicht auf dem Schulhof zurückgekommen. Wieder hat sie dazwischengehen müssen, als ein Schüler der siebten Klasse brutal auf einen am Boden liegenden Mitschüler einschlug. Grund: eine versehentlich umgestoßene Cola. Beide Schüler galten bisher als unauffällig. Mittlerweile erleben die Kolleginnen und Kollegen solche Szenen beinahe täglich. Wie kommt es zu der zunehmenden Gewalt? Frau B macht die Gewaltdarstellungen im Fernsehen verantwortlich, Herr A die familiären Verhältnisse.

Frau B: *Die Gewalt hat nicht nur zugenommen, sondern auch eine ganz neue Qualität bekommen! Heute genügt der kleinste Anlaß, dann wird zugeschlagen, viel brutaler als früher und auch dann noch, wenn der andere längst k.o. ist. Eigentlich wundert mich das gar nicht. Bei den Filmen, die die Schüler/innen sich heutzutage anschauen! Viele haben schon einen eigenen Fernseher und hocken stundenlang davor.*

Herr A: *Bis in die Nacht hinein. Um bloß keine Kindersendung zu verpassen! (1) - Nein, mir ist natürlich klar, daß die Kinder durch Fernsehen viel Gewalt mitkriegen, mehr als früher. Aber das ist meist keine reale Gewalt, und ich glaube, die Kinder wissen das. Viel schlimmer*

dürfte die Gewalt sein, die die Kinder zuhause erfahren, wenn etwa die Eltern zerstritten sind und sich gegenseitig fertigmachen.

Frau B: *Die reale Gewalt zuhause mag für Kinder schlimmer sein als die Gewaltdarstellungen im Fernsehen. Aber das betrifft nur wenige von denen, die hier in der Schule zuschlagen. Die meisten kommen aus intakten Familien. Ich denke, es liegt am Fernsehen. Dem kann sich keiner entziehen. Alle, die hier brutal werden, haben im Fernsehen schon mal brutale Szenen gesehen. Also muß ich damit rechnen, daß alle, die sich brutale Szenen im Fernsehen anschauen, später mal brutal werden. (3)*

Herr A: *Nein, da bin ich völlig anderer Meinung als Sie! (4) Es gibt viele, die sich brutale Sachen anschauen und trotzdem nicht gewalttätig werden. Viele Mädchen zum Beispiel, oder auch Jungen, die sich bei Gewaltszenen bloß dafür interessieren, mit welchen Tricks das filmtechnisch gemacht wird. Alles in allem glaube ich daher, daß die Gründe eher in der Familie liegen, selbst dann, wenn die Familie nach außen hin intakt ist. Dann ist es nämlich oft so: Da hat der Vater die Gewalt. Da hat die Mutter sich zu fügen. Da hat das Kind nicht aufzumucken. (2) Und in der Schule kommt der Frust dann raus.*

Auswahl von argumentativen Auffälligkeiten

Grundlage für eine inhaltsvalide Auswahl der *argumentativen* Auffälligkeiten war die oben dargestellte, mehrfach überprüfte und empirisch gesicherte Binnenstruktur des Konstrukts der «Argumentationsintegrität» ([Schreier, 1992](#); [Schreier & Groeben, 1990, 1995](#)). Dabei wurde inhaltliche Validität auf Merkmals-, Standard- und Strategienebene angestrebt. Den vier Merkmalen entsprechen Verstöße in den vier argumentativ auffälligen Szenarios einer Version. Realisiert wurden diejenigen Strategien, die nach der clusteranalytischen Auswertung der erwähnten Experten-Laien-Kategorisierungen als besonders typische Standardverletzungen des jeweiligen Merkmalsbereichs gelten können. Nach dem gleichen Prinzip wurden die Antwortalternativen für die Multiple-Choice-Vorgaben bestimmt. Diese enthalten neben der jeweils realisierten Strategie in jedem der Szenarios acht weitere Strategien, die jeweils zwei Strategien aus einem der vier Merkmale des Konstrukts der Argumentationsintegrität abdecken. Auf diese Weise enthält SPARK mehr als die Hälfte der ursprünglich 86 Strategien. Bei der Bearbeitung werden die Uptn zunächst gebeten anzugeben, ob sie eine Auffälligkeit bemerken und - falls dies zutrifft - in welcher Zeile sie diese Auffälligkeit lokalisieren. Anschließend ist unter neun Multiple-Choice-Vorgaben die zutreffende Auffälligkeit zu benennen. [Anhang A](#) zeigt beispielhaft die neun Antwortalternativen für die Angabe argumentativer Auffälligkeiten in Szenario L. Nach Angabe der entsprechenden Zeilennummer ist eine Auffälligkeit anzukreuzen; zwei weitere Auffälligkeiten können in freier Antwort benannt werden.

Auswahl rhetorischer Auffälligkeiten

Die Auswahl der *rhetorischen* Auffälligkeiten erfolgte in zwei Schritten. In einem ersten Schritt wurde aus Standardwerken zur Rhetorik ([Plett, 1991](#); [Ueding & Steinbrink, 1994](#)) ein Pool rhetorischer Stilfiguren gebildet. In die Auswahl gingen sowohl syntaktische als auch

semantische sowie pragmatische Auffälligkeiten ein, die unter dem Aspekt der persuasiven Wirksamkeit in besonderer Weise argumentationsrelevant erschienen (vgl. auch [Mischo et al., 1997](#)). Experten/innen beurteilten in einem zweiten Schritt die argumentative Relevanz der ausgewählten Stilfiguren. Im folgenden werden diejenigen zehn rhetorischen Stilfiguren aufgeführt und erläutert, die auf der Basis des Experten/innen-Ratings in die Szenarios eingearbeitet wurden: 1. Ironie («Du bist mir vielleicht ein schöner Freund.»); 2. rhetorische Frage («Wie lange willst Du noch warten?»); 3. Metapher («Stein des Anstoßes»); 4. Litotes («Das war keine leichte (statt: schwere) Aufgabe.»); 5. Anapher («Das ist gut. Das ist rein. Das ist uralte Lavendel.»); 6. Hyperbel (Übertreibung, z. B. «Riesenwaschkraft»); 7. Antithese («Some like it hot - some like it cold.»); 8. Oxymoron («schaurig schön», «schwarze Milch»); 9. syntaktische Umstellungen (z. B. Voranstellung wichtiger Satzteile); 10. Witz.

In sämtlichen Szenarios beider Versionen sind jeweils zwei der aufgeführten Auffälligkeiten enthalten. [Anhang B](#) zeigt am Beispiel von Szenario L das entsprechende Antwortfeld. Maximal zwei Auffälligkeiten sind zu identifizieren; zwei weitere können in freier Antwort benannt werden.

Generierung weiterer Argumente

Als weiterer Aspekt passiver argumentativ-rhetorischer Kompetenz bezieht sich passive argumentationslogische Kompetenz auf das Verständnis für die Logik einer Argumentation. Zur Erfassung dieses Kompetenzaspekts wurden für jedes Szenario weitere Argumente konstruiert, die die jeweils vorgetragene Position untermauern, unterminieren oder aber im betreffenden Diskussionskontext irrelevant sind. Die Uptn werden gebeten, sich den hypothetischen Fortgang der Argumentation vorzustellen. Sie lesen zunächst die strittige These, die den jeweiligen Stand der Diskussion zusammenfaßt. Anschließend besteht ihre Aufgabe darin, drei Argumente jeweils als Pro- oder Contra-Argument bzw. als ein in diesem Zusammenhang irrelevantes Argument einzustufen. Die Uptn werden darauf hingewiesen, daß alle drei Arten von Argumenten (Pro-, Contra-, irrelevant) vorkommen können, jedoch nicht vorkommen müssen. Als Maßstab für die Antwortrichtigkeit gelten die weitgehend übereinstimmenden Einschätzungen von Experten/innen aus einer Voruntersuchung (s. u.). [Anhang C](#) zeigt die für den hypothetischen Fortgang von Szenario L vorgegebene strittige These sowie eines der drei einzuschätzenden Argumente.

Realisierung von interaktiven Auffälligkeiten

Der Aspekt der passiven interaktiven Kompetenz bezieht sich zum einen auf die Identifikation von konfrontativen Äußerungen und zum anderen auf Aussagen, die zwischen den Beteiligten unstrittig sind. Als konfrontativ und gleichzeitig integer können nach den Ergebnissen der Voruntersuchung (s. u.) Äußerungen wie beispielsweise die folgenden gelten: «Nein! Da bin ich ganz anderer Auffassung als Sie!», «Diesen Standpunkt kann und will ich ganz und gar nicht teilen!» Bei den zu erkennenden gemeinsam geteilten Aussagen handelt es sich um szenariospezifische Abstraktionen vorgetragener Positionen, die nicht unmittelbar Gegenstand des betreffenden Gesprächs und insofern zu erschließen sind. Auch

in diesem Fall stellte die Voruntersuchung sicher, daß sich strittige von unstrittigen Aussagen weitgehend eindeutig unterscheiden lassen. [Anhang D](#) gibt die entsprechenden Antwortalternativen für Szenario L wieder.

Voruntersuchung

11 Experten/innen bearbeiteten eine vorläufige Fassung von Version A, 10 weitere Experten/innen eine vorläufige Fassung von Version B. An der Untersuchung nahmen solche Personen teil, die aufgrund ihrer beruflichen Qualifikation als argumentativ und rhetorisch kompetent eingeschätzt wurden, bis zu diesem Zeitpunkt jedoch nicht an der Entwicklung von SPARK beteiligt waren. Die Teilnehmer/innen bearbeiteten - analog zum Vorgehen in der späteren Hauptuntersuchung - beide Versionen von SPARK im Abstand von etwa zwei Wochen. Die meisten der Experten/innen fügten Kommentare an. Dabei handelte es sich im wesentlichen um Vorschläge zur Verbesserung von Formulierungen. Von besonderem Interesse für die Auswertung war die Übereinstimmung der Experten/innen-Antworten mit den als zutreffend angesetzten Antworten. Die Ergebnisse zeigen substantielle Übereinstimmungen hinsichtlich der argumentativen Auffälligkeiten ($\kappa = .87$ in Version A, $\kappa = .71$ in Version B), der rhetorischen Auffälligkeiten ($\kappa = .71$ bzw. $\kappa = .76$), der weiteren Argumente ($\kappa = .69$ bzw. $\kappa = .76$) sowie der (von den jeweiligen Sprechern/innen) gemeinsam geteilten Aussagen in Version B ($\kappa = .81$). Mäßige Übereinstimmungen resultierten für die (von den Sprecher/innen) gemeinsam geteilten Aussagen in Version A ($\kappa = .48$) sowie für die konfrontativen Äußerungen in Version A ($\kappa = .53$) und Version B ($\kappa = .43$). Die nähere Inspektion der Daten führte - unter Rückgriff auf die Vorschläge der Experten/innen - zu Formulierungsänderungen, die die Eindeutigkeit der zutreffenden Antworten weiter erhöhen sollten. Dies betraf insbesondere die (von den Sprecher/innen) gemeinsam geteilten Aussagen in Version A sowie die konfrontativen Äußerungen in beiden Versionen. Aus untersuchungstechnischen Gründen konnten die veränderten Versionen nicht durch weitere Experten/innen erneut überprüft werden. Allerdings läßt sich aufgrund der (im Sinne der Experten/innen) vorgenommenen Veränderungen vermuten, daß die angegebenen Übereinstimmungswerte eher als Untergrenze für die in der Validierungsstudie verwendete Fassung zu betrachten sind. Insgesamt können die als zutreffend angesetzten Antworten als ein durch Experten/innen-Konsens abgesichertes Kriterium passiver argumentativ-rhetorischer Kompetenz gelten.

Validierungsstudie

Untersuchungsplan, Teilnehmer/innen, Durchführung

SPARK wurde als kriteriumsorientierter Test vorrangig nach dem Kriterium der inhaltlichen Validität entwickelt. Die Ergebnisse dieses Vorgehens sollten jedoch auch nach Prinzipien der klassischen Testtheorie überprüft werden. Ziel war es, Werte für die üblichen Gütekriterien zu ermitteln und damit über das Kriterium der inhaltlichen Validität hinaus zusätzliche empirische Anhaltspunkte für die Validität von SPARK zu gewinnen. Eine Eichung

von SPARK war im Rahmen der betreffenden Untersuchung weder möglich noch beabsichtigt. Die Untersuchung erfolgte auf der Grundlage eines cross-over-Designs mit variiertem Bearbeitungsreihenfolge an einer größeren Stichprobe aus unterschiedlichen Alters-, Bildungs- und Berufsgruppen. Von insgesamt 144 Personen, die die Untersuchungsunterlagen (teilweise per Post) zur selbständigen Bearbeitung erhielten, nahmen 123 Untersuchungspartner/innen (Uptn) (69 Frauen und 54 Männer) zum ersten Erhebungszeitpunkt teil. Von diesen bearbeiteten 116 Uptn jeweils beide Versionen im Abstand von etwa vier Wochen. In Anbetracht der Schwierigkeit, genügend Uptn für die aufwendige Untersuchung zu gewinnen, wurde auf eine zweimalige Vorgabe der gleichen Version verzichtet. Die entsprechenden Werte für die Retest-Reliabilität sollen in einer geplanten zukünftigen Validierungsstudie mit erhoben werden.

Die Bearbeitung begann für alle Uptn mit einer einführenden Übersicht über die folgenden Aufgaben. Ein auf das Rahmenthema der jeweiligen Testversion zugeschnittenes Beispiel verdeutlichte die Grundstruktur eines Szenarios und erläuterte den Modus der Aufgabenbeantwortung. Anschließend folgte in Version A eine kurze inhaltliche Einführung in das Rahmenthema «Macht Fernsehen aggressiv?», die sich in Version B entsprechend auf das Rahmenthema «Ist Intelligenz angeboren?» bezog. Nach dieser Einführung wurden in beiden Versionen die fünf Szenarios in zufälliger Position und Reihenfolge präsentiert. Abschließend wurden die Uptn um Angaben zu ihrer Person gebeten. [Abbildung 1](#) gibt eine Übersicht über den Ablauf der Bearbeitung.

Ergebnisse

Gesamtwerte

Von 180 möglichen Punkten für argumentativ-rhetorische Kompetenz erzielten die Uptn in Version A durchschnittlich 110.8 Punkte, in Version B 98.0 Punkte. Uptn mit höheren Bildungsabschlüssen profitierten von ihren Kenntnissen in erwartbarer Weise: Die Korrelation zwischen den Gesamtpunktwerten für passive argumentativ-rhetorische Kompetenz und den fünf unterschiedenen Bildungsstufen (abgeschlossenes Studium, Abitur, Fachabitur, Realschule, Hauptschule) beträgt für Version A .27, für Version B .30. Uptn mit systematischen argumentativ-rhetorischen Vorkenntnissen ($N = 20$) erzielten in Version B mit durchschnittlich 104.8 Punkten erwartungsgemäß höhere Punktwerte für ihre passive argumentativ-rhetorische Kompetenz als Uptn, die nicht über systematische argumentativ-rhetorische Vorkenntnisse verfügten ($N = 89$) und eine mittlere Punktzahl von 97.0 Punkten erreichten. Geschlechtsspezifische Unterschiede hinsichtlich der passiven argumentativ-rhetorischen Kompetenz traten nicht auf: Die entsprechenden Mittelwerte und Standardabweichungen sind nahezu identisch, wie aus dem zusammenfassenden Überblick über die wichtigsten statistischen Kennwerte in [Tabelle 2](#) hervorgeht.

[Tabelle 2](#) zeigt zum einen die insgesamt höhere Schwierigkeit von Version B, die in [Tabelle 4](#) auf Szenarioebene weiter verdeutlicht wird. Um Ergebnisse aus den beiden unterschiedlich schwierigen Versionen vergleichen zu können, wurden inzwischen Tabellen nach der Äquiperzentil-Methode erstellt. Die sehr niedrigen Minimumwerte in beiden

Versionen lassen sich als Hinweis auf einzelne nachlässige Bearbeitungen interpretieren. Hinsichtlich des Vorwissens der Uptn wird ein Befund von [Blickle \(1993\)](#) bestätigt, der einen positiven Zusammenhang von Schulbildung und der korrekten Diagnose argumentativer Unintegrität fand. Von ihren argumentativ-rhetorischen Vorkenntnisse profitierten die Uptn vor allem in der schwierigeren Version B. Dies kann als Beleg für die Validität des Konstrukts gewertet werden. Für die geschlechtsspezifische Fairness von SPARK spricht die inzwischen replizierte Beobachtung, daß bezüglich der passiven argumentativ-rhetorischen Kompetenz keine bedeutsamen geschlechtsspezifischen Mittelwertsunterschiede auftreten ([Flender, 1997](#); [Flender & Christmann, 1999](#)).

Korrelation der Versionen und Dimensionen

Version A und Version B korrelieren insgesamt zu .75. Dabei besteht hinsichtlich der Dimension 1 (argumentative Aspekte) ein Zusammenhang von .39, bezüglich der Dimension 2 (rhetorische Aspekte) eine Korrelation von .69. Die Dimensionen 3 (argumentationslogische Aspekte) beider Versionen korrelieren zu .56; für Dimension 4 (interaktive Aspekte) ergibt sich ein Wert von .37. Der Zusammenhang von .39 bezüglich der Dimension 1 beider Versionen fällt möglicherweise nicht höher aus, weil in den Aufgaben zu den argumentativen Aspekten in Version A andere Teilaspekte des Konstrukts der Argumentationsintegrität (Strategien unintegren Argumentierens) realisiert wurden als in Version B. Der eher niedrige Zusammenhang von .37 zwischen den beiden Realisierungen von Dimension 4 läßt sich vermutlich darauf zurückführen, daß die konfrontativen Äußerungen auch nach den im Anschluß an die Voruntersuchung vorgenommenen Änderungen nicht in beiden Versionen gleichermaßen eindeutig lokalisierbar sind.

Die Korrelation von .75 zwischen beiden Versionen läßt sich nur unter Berücksichtigung weiterer Resultate als Reliabilitätsmaß interpretieren, da ein Einfluß der Bearbeitungsreihenfolge vorliegt: Die Versionen korrelieren zu .83, wenn Version B vor Version A bearbeitet wurde. Bei umgekehrter Bearbeitungsreihenfolge liegt der Zusammenhang bei .53. Allerdings erzielten die Uptn in diesem Fall durchschnittlich höhere Gesamtwerte. Diese Effekte bedürfen einer weitergehenden Analyse in zukünftigen Untersuchungen. Vorläufig läßt sich vermuten, daß sich die höheren Durchschnittswerte in der Bearbeitungsreihenfolge A-B zum Teil möglicherweise auf didaktische und motivationale Vorteile von Version A zurückführen lassen. Motivierend könnte dabei die tendenziell geringere Schwierigkeit von Version A (vgl. [Tab. 4](#)) wirken, die ihrerseits eventuell aus der Rahmenthematik sowie deren Realisierung resultiert. Die genannten Unterschiede verdeutlichen, daß der in den geschilderten Konstruktionsprinzipien begründeten inhaltlichen Parallelität der beiden Versionen nur teilweise eine statistische Parallelität (Äquivalenz) der Testformen entspricht. [Tabelle 3](#) gibt eine zusammenfassende Übersicht über die Interkorrelationen der Dimensionen in beiden Versionen.

Schwierigkeit

SPARK ist als ein individueller Kompetenztest konzipiert worden, der auf unterschiedlichen

Schwierigkeitsniveaus differenzieren und zugleich Spielraum für die Erfassung von Trainingseffekten enthalten soll. In Übereinstimmung mit dieser Zielsetzung enthält SPARK unterschiedlich schwierige Aufgaben auf einem insgesamt hohen Schwierigkeitsniveau, wie [Tabelle 4](#) verdeutlicht.

Wie aus [Tabelle 4](#) hervorgeht, gelang es nur wenigen Uptn, zutreffende Antworten in sämtlichen Dimensionen abzugeben. Für die überraschend deutlichen Unterschiede in der Schwierigkeit der einzelnen Szenarios sind mehrere Interpretationsmöglichkeiten diskutiert worden (vgl. [Flender et al., 1996](#)). Dabei wurde beispielsweise hinsichtlich der argumentativen Aspekte (Dimension 1) vermutet, daß einige Strategien unfairen Argumentierens (z. B. «Diskreditieren», Szenario S) generell leichter erkannt werden als andere Strategien (z. B. «falsche Behauptung aufstellen», Szenario E). Besonders niedrige Lösungswahrscheinlichkeiten könnten prinzipiell auch auf Mängel bei der Aufgabengestaltung verweisen. Dies erscheint jedoch angesichts der insgesamt hohen Übereinstimmung der Experten/innen (s. o.) eher unwahrscheinlich. Eine Ausnahme bildet dabei Szenario FM. Bei diesem auch von den Experten/innen uneinheitlich bearbeiteten Szenario haben die im Anschluß an die Voruntersuchung vorgenommenen Formulierungsänderungen offenbar nicht ausgereicht, um die Eindeutigkeit der Antworten sicherzustellen. Die übrigen Szenarios (einschließlich der Szenarios mit hohen Schwierigkeitswerten) wurden von den Experten/innen überwiegend mit hoher Übereinstimmung bearbeitet. Aus diesem Grund lassen sich hohe Schwierigkeiten in den meisten Fällen als Spielraum zur Erfassung von Trainingseffekten interpretieren. Ein Deckeneffekt ist dabei allenfalls in Dimension 3 (weitere Argumente) in Kauf zu nehmen.

Homogenität, Trennschärfe, interne Konsistenz

Als Maß für die Homogenität bzw. Heterogenität eines Tests gelten die Interkorrelationen der Dimensionen (vgl. [Lienert & Raatz, 1994](#), S. 99 f.). Während die vier Dimensionen in Version A zwischen .21 und .51 interkorrelieren, liegen die entsprechenden Werte in Version B im Bereich von .23 bis .48. Dabei bestehen in beiden Versionen die höchsten Zusammenhänge zwischen Dimension 2 (rhetorische Auffälligkeiten) und Dimension 3 (weitere Argumente). Diese beiden Dimensionen weisen darüber hinaus in beiden Versionen eine höhere Trennschärfe auf als die Dimensionen 1 und 4, wie aus den in [Tabelle 5](#) wiedergegebenen Korrelationen mit dem Gesamtwert der jeweiligen Version zu ersehen ist. Insgesamt resultiert in Version A ein Wert von .68 für Cronbachs α . In Version B beträgt der entsprechende Wert .71.

Individuelle Urteilssicherheit

Bezüglich der individuellen Urteilssicherheit ist zwischen richtigen und falschen Antworten unterschieden worden (s. o. «Konstruktexplikation»). Für richtige Antworten wurde dabei ein positiver, für falsche Antworten ein negativer Zusammenhang mit passiver argumentativ-rhetorischer Kompetenz angenommen: Personen mit hohen Gesamtwerten für argumentativ-rhetorische Kompetenz sollten bei richtigen Antworten sicherer, bei falschen Antworten unsicherer urteilen als Personen mit niedrigen Gesamtwerten für passive

argumentativ-rhetorische Kompetenz. Unter dem Vorbehalt ihrer empirischen Bewährung gingen diese Annahmen in die Konstruktextplikation sowie in das vorläufige Punktevergabemodell ein. Die Resultate zeigen, daß richtige Antworten - unabhängig von der passiven argumentativ-rhetorischen Kompetenz der Uptn - generell mit höherer Urteilssicherheit gegeben wurden als falsche Antworten (Version A: $F = 111.67$; $p = .0001$; Version B: $F = 136.33$; $p = .0001$). Entgegen unserer Annahmen antworteten Uptn mit hohen Werten für passive argumentativ-rhetorische Kompetenz bei richtigen Antworten nicht mit höherer Sicherheit (bzw. bei falschen Antworten nicht mit geringerer Sicherheit) als Uptn mit niedrigen Werten für passive argumentativ-rhetorische Kompetenz. Bei den betreffenden Korrelationen zwischen den Gesamtwerten für passive argumentativ-rhetorische Kompetenz und den durchschnittlichen individuellen Werten für die Urteilssicherheit bei richtigen bzw. falschen Antworten handelt es sich nahezu ausnahmslos um Nullkorrelationen. Eine neuere Untersuchung ([Flender, 1997](#)) führte bezüglich der Urteilssicherheit zu ähnlichen Resultaten. Damit konnte die differenzierte Betrachtung der Urteilssicherheit als Teil des Konstrukts der passiven argumentativ-rhetorischen Kompetenz empirisch nicht bestätigt werden. Dies wird in zukünftigen Untersuchungen (und einer ggf. damit verbundenen Revision von SPARK) zu berücksichtigen sein. Nennenswerte Änderungen bezüglich der berichteten empirischen Kennwerte sind für den Fall, daß auf die Einbeziehung der Urteilssicherheit verzichtet wird, nicht zu erwarten, wie eine entsprechende Analyse der Daten zeigte.

Gesamtdiskussion

Das Konstrukt der passiven argumentativ-rhetorischen Kompetenz bezieht sich auf die individuelle Diagnosefähigkeit bezüglich argumentativer, rhetorischer, argumentationslogischer sowie interaktiver Aspekte von Argumentationen. Zur Erfassung dieses Konstrukts ist SPARK als kriteriumsorientierter Test unter dem vorrangigen Aspekt der inhaltlichen Validität entwickelt und in einer Voruntersuchung erprobt worden. Mit der Validierungsstudie sollten die beiden inhaltlich parallelen Versionen an einer größeren Stichprobe erprobt und die Daten nach Prinzipien der klassischen Testtheorie ausgewertet werden. Die Ergebnisse dieser Studie stehen in Einklang mit der inhaltlichen Vielschichtigkeit des Konstrukts der passiven argumentativ-rhetorischen Kompetenz. Für die Validität des Konstrukts spricht der Zusammenhang zwischen passiver argumentativ-rhetorischer Kompetenz und (Schul-)Bildung (.27 in Version A; .30 in Version B), der bereits bei [Blickle \(1993\)](#) festgestellt wurde. Dieses Resultat wird zusätzlich durch die Beobachtung gestützt, daß Uptn mit systematischen argumentativ-rhetorischen Vorkenntnissen in der schwierigeren Version B höhere Werte für ihre passive argumentativ-rhetorische Kompetenz aufweisen als Uptn ohne derartige Vorkenntnisse, während Alter und Beruf in diesem Zusammenhang ohne Bedeutung sind. Hinsichtlich der Reliabilität sind die Werte uneinheitlich. Die inhaltlich parallelen Versionen korrelieren im Mittel zu .75. Cronbachs α beträgt in Version A .68, in Version B .71. Demgegenüber liegen die Werte für Homogenität, Trennschärfe und interne Konsistenz auf Dimensions- und Szenarioebene auf einem insgesamt niedrigen bis mäßigen Niveau. Dies läßt sich unter anderem auf deutliche

Schwierigkeitsunterschiede sowie auf Einflüsse aufgrund der Reihenfolge der bearbeiteten Versionen zurückführen. Dabei bedürfen die Effekte aufgrund der Bearbeitungsreihenfolge einer weiteren Klärung. Die deutlichen Schwierigkeitsunterschiede lassen sich nur vereinzelt auf verbliebene Mängel in der Operationalisierung zurückführen. Betrachtet man die Übereinstimmung der Experten/innen aus der Voruntersuchung als ausreichenden Maßstab für die Richtigkeit und Eindeutigkeit der Antworten, dann lassen sich die unterschiedlichen Schwierigkeiten als notwendiges Merkmal eines Niveautests interpretieren, der Spielräume für Trainingsmaßnahmen spezifizieren und die Effektivität derartiger Maßnahmen erfassen soll. Eine erneute Überprüfung der testtheoretischen Kennwerte (möglicherweise verbunden mit einer Revision von SPARK) soll in einer geplanten zukünftigen Untersuchung erfolgen. Dabei kann nach den bisher vorliegenden Ergebnissen das Geschlecht als relevanter Faktor für passive argumentativ-rhetorische Kompetenz vernachlässigt werden: Geschlechtsspezifische Unterschiede bezüglich der passiven argumentativ-rhetorischen Kompetenz fanden sich weder in dieser, noch in einer neueren Untersuchung ([Flender & Christmann, 1999](#)). Dies läßt sich auch als Beleg für die Fairneß des Tests interpretieren. Vernachlässigbar ist nach den bisherigen empirischen Befunden auch die Urteilssicherheit der Uptn, die als (empirisch zu überprüfender) Teil des Konstrukts der passiven argumentativ-rhetorischen Kompetenz nicht bestätigt werden konnte. Die Durchführungs- und Auswertungsökonomie ist inzwischen durch eine computergestützte Fassung (WinSPARK; [Flender, 1997](#); [Flender & Christmann, 1999](#)) weiter verbessert worden. [1]

SPARK kann vollständig oder in Teilen zu einem oder mehreren aufeinanderfolgenden Erhebungszeitpunkten verwendet werden. Als Forschungsinstrument ist SPARK bereits erfolgreich eingesetzt worden ([Sladek, Christmann & Groeben, 1996](#)). Der Einsatz als Trainingsevaluationsinstrument ist für zukünftige Schritte der praktischen Anwendung des Konstrukts «Argumentationsintegrität» vorgesehen.

Literatur

Blickle, G. (1993). *Argumentationsintegrität (IX): Personale Antezedenzbedingungen der Diagnose argumentativer Unintegrität*. Arbeiten aus dem Sonderforschungsbereich 245, Bericht Nr. 57, Heidelberg/Mannheim.

Christmann, U., Schreier, M. & Groeben, N. (1996). War das Absicht? Indikatoren subjektiver Intentionalitätszustände bei der ethischen Bewertung von Argumentationsbeiträgen. *Linguistik und Literaturwissenschaft, 101*, (Themenheft «Sprache und Subjektivität I»). 70-113.

Flender, J. (1997). *Entwicklung und Validierung einer computerunterstützten Ökonomieversion der Skala der passiven argumentativ-rhetorischen Kompetenz (Win-SPARK)*. Unveröff. Diplomarbeit, Universität Heidelberg.

Flender, J. & Christmann, U. (1999). Entwicklung einer computerunterstützten Ökonomieversion der Skala der passiven argumentativ-rhetorischen Kompetenz. *Medienpsychologie, 11*, 21-37.

- Flender, J., Christmann, U., Groeben, N. & Mlynski, G. (1996). *Argumentationsintegrität (XVIII): Entwicklung und erste Validierung einer Skala zur Erfassung der passiven argumentativ-rhetorischen Kompetenz (SPARK)*. Arbeiten aus dem Sonderforschungsbereich 245, Bericht Nr. 97, Heidelberg/Mannheim.
- Groeben, N., Christmann, U. & Mischo, C. (1997). Die Entwicklung eines Trainings zum Umgang mit unintegrem Argumentieren. Teil I: Forschungsbasis, Grundkonzeption und Dimensionen des Trainings. *Sprechen Zeitschrift für Sprechwissenschaft*, 15, 59-76.
- Groeben, N., Schreier, M. & Christmann, U. (1993). Fairness beim Argumentieren: Argumentationsintegrität als Wertkonzept einer Ethik der Kommunikation. *Linguistische Berichte*, 147, 355-382.
- Hofer, M., Fleischmann, T. & Pikowsky, B. (1991). *Jugendliche unterschiedlichen Alters im argumentativen Konfliktgespräch mit ihrer Mutter*. Arbeiten aus dem Sonderforschungsbereich 245, Bericht Nr. 36, Heidelberg/Mannheim.
- Infante, D. A. & Rancer, A. S. (1982). A conceptualization and measure of argumentativeness. *Journal of Personality Assessment*, 46, 72-80.
- Koehler, D. (1992). *Die Entwicklung der Argumentationsfähigkeit von Kindern*. Unveröffentlichte Dissertation, Universität Mainz.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Lienert, G. A. & Raatz, U. (1994). *Testaufbau und Testanalyse* (5., völlig Neubearb. u. erw. Aufl.). Weinheim: Beltz, Psychologie Verlags Union.
- Miller, M. (1980). Zur Ontogenese moralischer Argumentationen. *Zeitschrift für Literaturwissenschaft und Linguistik*, 38/39, 58-108.
- Miller, M. (1982). Argumentationen als moralische Lernprozesse. *Zeitschrift für Pädagogik*, 28, 299-314.
- Miller, M. (1984). Zur Ontogenese des koordinierten Dissens. In W. Edelstein & J. Habermas (Hrsg.), *Soziale Interaktion und soziales Verstehen* (S. 220-250). Frankfurt/M.: Suhrkamp.
- Mischo, C., Groeben, N. & Christmann, U. (1997). Was ist in Argumentationen überzeugender: rhetorische Ästhetik oder kommunikative Fairneß?. *Zeitschrift für Experimentelle Psychologie*, 44, 656-685.
- Mischo, C., Flender, J., Christmann, U. & Groeben, N. (1999). Argumentational integrity: A training program for dealing with unfair argumentational contributions. In F.H. van Eemeren, R. Grootendorst, J.A. Blair & C.A. Willard (Eds.), *Proceedings of the Fourth ISSA Conference on Argumentation, Amsterdam 1998* (pp. 586-590). Amsterdam: Sic Sat.
- Overton, W. F., Ed. (1990). *Reasoning, necessity, and logic: Developmental perspectives*. Hillsdale, NJ: Erlbaum.
- Plett, H. F. (1991). *Einführung in die rhetorische Textanalyse* (8. Aufl.). Hamburg: Buske.
- Rancer, A. S., Kosberg, R. L. & Baukus, R. A. (1992). Beliefs about arguing as predictors of trait argumentativeness: Implications for training in argument and conflict management. *Communication Education*, 41, 376-387.

- Schreier, M. (1992). *Rhetorische Strategien und Integritätsstandards: Zur Relation von Rhetorik, Dialektik und Argumentationsintegrität*. Unveröffentlichte Diplomarbeit, Universität Heidelberg.
- Schreier, M. (1994). Unfaire Strategien der Gebrauchsrhetorik: Im Spannungsfeld zwischen Norm und Normalität. In B. Spillner (Hrsg.), *Fachkommunikation. Kongreßbeiträge der 24. Jahrestagung der Gesellschaft für Angewandte Linguistik*. GAL e.V. (S. 229 f.). Frankfurt/M: Lang.
- Schreier, M. (1997). *Das Erkennen sprachlicher Täuschung. Über Absichtlichkeitsindikatoren beim unintegren Argumentieren*. Münster: Aschendorff.
- Schreier, M. & Groeben, N. (1990). *Argumentationsintegrität (III): Rhetorische Strategien und Integritätsstandards*. Arbeiten aus dem Sonderforschungsbereich 245, Bericht Nr. 30, Heidelberg/Mannheim.
- Schreier, M. & Groeben, N. (1995a). Unfairness in argumentative discussions: Relevant factors and prototypical manifestations in spoken language. In F. H. van Eemeren, R. Grootendorst, A. J. Blair & C. A. Willard (Eds.), *Proceedings of the Third ISSA Conference on Argumentation Amsterdam 1994, Vol. III* (pp. 276-286). Amsterdam: Sic Sat.
- Schreier, M. & Groeben, N. (1996). Ethical guidelines for the conduct in argumentative discussions: An explanatory study. *Human Relations*, 49, 123-132.
- Schreier, M., Groeben, N. & Blickle, G. (1995). The effects of (un-)fairness and (im-) politeness on the evaluation of argumentative communication. *Journal of Language and Social Psychology*, 14, 260-288.
- Schreier, M., Groeben, N. & Christmann, U. (1995). «That's not fair!» Argumentational integrity as an ethics of argumentative communication. *Argumentation*, 9, 267-289.
- Sladek, U., Christmann, U. & Groeben, N. (1996). *Argumentationsintegrität (XVI): Der Einfluß personaler und interaktiver Bedingungen auf die Bewertung argumentativer (Un-) Integrität*. Arbeiten aus dem Sonderforschungsbereich 245, Bericht Nr. 95, Heidelberg/Mannheim.
- Stein, N. L. & Miller, C. A. (1989). I win - you lose: The development of argumentative thinking. In J. F. Voss, D. N. Perkins & J. Segal (Eds.), *Informal reasoning and instruction* (pp. 265-290). Hillsdale, NJ: Erlbaum.
- Stein, N. L. & Miller, C. A. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology*, (Vol. 4, pp. 285-335). Hillsdale, NJ: Erlbaum.
- Ueding, G. & Steinbrink, B. (1994). *Grundriß der Rhetorik: Geschichte, Technik, Methoden* (3., überarb. u. erw. Aufl.). Stuttgart: Metzler.
- Völzing, P.-L. (1982). *Kinder argumentieren: Die Ontogenese argumentativer Fähigkeiten* (Informationen zur Sprach- und Literaturdidaktik, 36). Paderborn: Schoeningh.

Anhang

a.

Multiple-Choice-Vorgaben argumentativer Auffälligkeiten in Szenario L (die dritte Antwort trifft zu; vgl. das zugehörige Beispielgespräch unter Testkonstruktion).

ganz ganz
sicher unsicher

- | | |
|--|----------------------|
| <input type="checkbox"/> Eine Person hält sich ein Hintertürchen offen. Erläuterung: Jemand stellt seine/ihre Überzeugungen nicht klar dar, sondern deutet sie nur schwammig an. Später kann er/sie behaupten: 'Das habe ich so nicht gesagt.' | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person stellt gegenüber der anderen Person eine unerfüllbare Forderung auf. Erläuterung: Eine Person fordert eine andere zu etwas auf, wozu diese unmöglich in der Lage ist. - Beispiel: Ein Vater sagt zu seinem 15jährigen Sohn, der noch zur Schule geht: "Zieh doch aus, wenn es Dir hier nicht paßt!" | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person macht einen unzulässigen Umkehrschluß. Erläuterndes Beispiel: "Wer Leistungssport treibt, ernährt sich gesund. Daraus folgt: Wer sich gesund ernährt, treibt Leistungssport." | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person verallgemeinert in unzulässiger Weise. Erläuterung: Von einem Einzelbeispiel wird auf ein allgemeines Prinzip geschlossen. - Beispiel: "Einige Motorradfahrer rasen. Also sind alle Motorradfahrer Raser." | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person schüchtert die andere Person durch Grobheiten ein. | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person stützt ihre Position ausschließlich durch Appell an die Gefühle einer anderen Person. Erläuterndes Beispiel: "Ich bin für Evolution, nicht für Revolution!" | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person versteht die andere Person falsch. Erläuterung: Eine Person gibt dem, was eine andere Person gesagt hat, einen anderen Sinn, z.B. indem sie einen bildhaften Vergleich wörtlich nimmt. | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person bezweifelt unberechtigt die moralische Redlichkeit der anderen Person. Erläuterung: Statt sachlich zu argumentieren, versucht eine Person <i>unberechtigt</i> , die moralische Glaubwürdigkeit der anderen Person zu beschädigen. - Beispiel: "Ein bißchen mehr Ehrlichkeit hätte ich schon von Ihnen erwartet!" | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Eine Person unterbricht die andere Person zum wiederholten Mal. | 1 — 2 —
3 — 4 — 5 |
| <input type="checkbox"/> Keine der genannten Auffälligkeiten | 1 — 2 —
3 — 4 — 5 |

b.

Multiple-Choice-Vorgaben rhetorischer Auffälligkeiten in Szenario L (die zweite und die fünfte Antwort treffen zu; vgl. das zugehörige Beispielgespräch unter Testkonstruktion).

ganz ganz
sicher unsicher

- In / ab Zeile: _____ benutzt eine Person eine Metapher.** Erläuternde Beispiele: "Fuß des Berges", "Stein des Anstoßes", "Baulöwe", "geistige Brandstifter". 1 — 2
— 3 —
4 — 5
- In / ab Zeile: _____ drückt sich eine Person ironisch aus.** Erläuterung: Ironie liegt vor, wenn eine Person etwas anderes sagt als sie eigentlich meint. Beispiel: "Sie haben mir ja eine schöne Bescherung bereitet!" 1 — 2
— 3 —
4 — 5
- In / ab Zeile: _____ untertreibt eine Person.** Erläuterndes Beispiel: Ein Popstar hat in einer ausverkauften Konzerthalle die Jugendlichen zu Begeisterungstürmen hingerissen. Im anschließenden Interview meint er: "Nun, es scheint den Leuten ein bißchen gefallen zu haben." 1 — 2
— 3 —
4 — 5
- In / ab Zeile: _____ drückt sich eine Person auffällig aus, indem sie mehrere aufeinanderfolgende Sätze mit den gleichen Worten beginnt.** Beispiel: "Das ist gut. Das ist rein. Das ist uralte Lavendel." 1 — 2
— 3 —
4 — 5
- In / ab Zeile: _____ drückt sich eine Person auffällig aus, indem sie mehrere aufeinanderfolgende Sätze mit den gleichen Worten beendet.** Beispiel: "Arbeit ist wichtig. Freizeit ist wichtig. Gesundheit ist wichtig." 1 — 2
— 3 —
4 — 5

c.

Aufgabe zur Erfassung der passiven argumentationslogischen Kompetenz am Beispiel von Szenario L: strittige These und erstes von insgesamt drei einzuschätzenden Argumenten (die zweite Antwort trifft zu; vgl. das zugehörige Beispielgespräch unter Testkonstruktion).

Strittige These: "Es liegt an den Gewaltdarstellungen im Fernsehen, daß die Gewalt unter den Schülern/innen in der Schule zugenommen hat."

Argument 1:

"In der Zeitung habe ich neulich von einer großen Studie zur Gewalt im Fernsehen gelesen. Da stand sinngemäß: 'Durch das Anschauen von aggressionsgeladenen Szenen kann die Gewaltbereitschaft der Zuschauer/innen verringert werden. Brutale Verhaltensweisen sind als Folge der gesehenen Szenen nicht zu erwarten.' (Mannheimer Morgen, 30.3.1995)."

Dieses Argument kann im vorliegenden Zusammenhang *am ehesten* verwendet werden als

ganz ganz
sicher unsicher

- | | |
|---|-------------------------|
| <input type="checkbox"/> PRO - Argument ("Ja, es liegt an den Gewaltdarstellungen im Fernsehen, daß die Gewalt unter Schülern/innen in der Schule zugenommen hat.") | 1 — 2
— 3 —
4 — 5 |
| <input type="checkbox"/> CONTRA - Argument ("Nein, es liegt nicht an den Gewaltdarstellungen im Fernsehen, daß die Gewalt unter Schülern/innen in der Schule zugenommen hat.") | 1 — 2
— 3 —
4 — 5 |
| <input type="checkbox"/> weder PRO noch CONTRA - Argument (Argument ist für <i>diese</i> These bedeutungslos.) | 1 — 2
— 3 —
4 — 5 |

d.

Multiple-Choice-Vorgaben für die gemeinsam geteilte Aussage in Szenario L (die zweite Antwort trifft zu; vgl. das zugehörige Beispielgespräch unter Testkonstruktion).

ganz ganz
sicher unsicher

- | | |
|---|-------------------------|
| <input type="checkbox"/> "Die meisten der gewalttätigen Kinder kommen aus zerstrittenen Familien." | 1 — 2
— 3 —
4 — 5 |
| <input type="checkbox"/> "Kinder sitzen lange vorm Fernseher und bekommen auf diese Weise viel Gewalt mit." | 1 — 2
— 3 —
4 — 5 |
| <input type="checkbox"/> "Gewaltdarstellungen im Fernsehen verleiten die meisten Kinder zur Nachahmung dieser Gewalt." | 1 — 2
— 3 —
4 — 5 |
| <input type="checkbox"/> keine der genannten Aussagen | 1 — 2
— 3 —
4 — 5 |

Fussnoten

1. WinSPARK beschränkt sich im Unterschied zu SPARK auf die Erfassung der passiven argumentativen Kompetenz. Mit Blick auf eine Verbesserung der internen und externen Validität sieht WinSPARK vor, daß argumentative Regelverletzungen - ähnlich wie in mündlicher Alltagskommunikation - innerhalb einer bestimmten Zeit und ohne die Möglichkeit einer späteren Korrektur identifiziert werden müssen. Ein Überspringen von Antworten ist nicht möglich. Sämtliche Antworten werden automatisch erfaßt und einschließlich der berechneten Punktzahlen und Bearbeitungszeiten in einer gesonderten Datenbank abgelegt. In einer Äquivalenzstudie ($N = 63$) zeigten sich hohe Übereinstimmungen hinsichtlich der Mittelwerte und Standardabweichungen von Papier-Bleistift- und Computerfassung. Die mittleren Korrelationen von Papier-Bleistift- und Computerfassung betragen für Version A .48, für Version B .61.

Anschrift

Jürgen Flender, Psychologisches Institut, Universität Heidelberg, Hauptstraße 47-51, D-69117 Heidelberg, Tel: +49 6221 547308, Fax: +49 6221 547273, Email: Juergen_Flender@psi-sv2.psi.uni-heidelberg.de.

Tabellen

1. Bedingungen, Merkmale, Standards und ausgewählte Strategien (un-)integren Argumentierens; ausgewählte Standards kursiv; Strategien mit Angabe der zugehörigen Version (A bzw. B)
 2. Statistische Kennwerte für die Versionen A und B
 3. Interkorrelationen der Dimensionen in den Versionen A und B
 4. Schwierigkeitswerte der Szenarios, Dimensionen und Versionen, angegeben in Prozent richtiger Antworten
 5. Korrelationen zwischen den Dimensionen und dem Gesamtwert (GW) unter Ausschluß der jeweiligen Dimension aus dem Gesamtwert; Cronbachs α für die betreffende Version unter Ausschluß der jeweiligen Dimension aus dem Gesamtwert sowie für die betreffende Version insgesamt; Angaben für beide Versionen
-

Abbildungen

1. Ablauf der Bearbeitung von SPARK