

„Lautes Denken“ als qualitative Methode zur Untersuchung der Validität von Testitems

Erkenntnisse einer Studie zur Diagnose des ökonomischen Fachwissens von Schülerinnen und Schülern der Sekundarstufe I

Christina Schnell

Institut für Ökonomische Bildung an der Carl von Ossietzky Universität Oldenburg

Zusammenfassung

In der ökonomischen Bildung werden zunehmend mehr standardisierte Testverfahren entwickelt, anhand derer ökonomische Kompetenzen und Fachwissen gemessen werden können. Die Erhebungsinstrumente müssen dabei hinreichend valide und reliabel sein, damit diese aus wissenschaftlicher Perspektive eingesetzt werden dürfen. In diesem Beitrag wird die Methode „Lautes Denken“ als eine Möglichkeit vorgestellt, mithilfe derer die Validität von Testaufgaben untersucht werden kann. Durch die Verbalisierung der Gedanken während der Aufgabenbearbeitung kann sichtbar gemacht werden, wie die Testaufgaben verstanden werden und wie Schülerinnen und Schüler ihre Lösung aus den vorgegebenen Antworten auswählen. Die Methode „Lautes Denken“ wurde im Rahmen einer Studie zur Erfassung ökonomischen Fachwissens von Schülerinnen und Schülern der Schuljahrgänge acht und neun an Haupt-, Real- und Oberschulen des Sekundarbereichs I angewendet. Mithilfe dieser explorativen Studie konnten Strategien von Schülerinnen und Schülern zur Auswahl ihrer Lösungen bei gebundenen Testaufgaben identifiziert und die konzipierten Items hinsichtlich ihrer Validität untersucht werden.

Abstract

Economic Education is experiencing the growth of more standardized tests, through which economic competencies and expertise can be measured. The survey instruments have to be sufficiently valid and reliable so that they can be used for scientific investigation. This paper presents the method of “loud thinking” as a possibility, with whose help the validity of test tasks can be examined. The question of how test tasks are understood and how students select their solutions from the answers provided can be made visible by verbalizing thoughts during the process of creating the tasks. The method of “loud thinking” was used in the context of a study to capture economic understanding of students from year eight and nine in the following German secondary schools: *Hauptschule*, *Realschule* and *Oberschule*. Using this explorative study, strategies adopted by students to select answers for bound test tasks can be identified and the conceived test items can be examined for their validity.

1 Einleitung

In der ökonomischen Bildung werden – wie in anderen Fachdidaktiken auch – zunehmend standardisierte Tests entwickelt und erprobt, mit denen der Stand des Wirtschaftswissens oder die Ausprägung ökonomischer (Teil-)Kompetenzen erhoben werden soll. Aus wissenschaftlicher Sicht müssen solche Testinstrumente reliabel und valide sein. Ob ein Test das misst, was er auch messen soll, kann abhängig von der Zielsetzung des Tests, durch verschiedene Formen der Validität (curriculare Validität, Konstrukt-, Inhalts-, und/oder Kriteriumsvalidität) erfasst werden. Die Validität kann wiederum mit unterschiedlichen Methoden und anhand verschiedener Kriterien geprüft werden (vgl. Hartig et al. 2012, 144). Der vorliegende Beitrag fokussiert die Methode „Lautes Denken“, mit deren Hilfe die Validität eines Tests auf der Ebene der einzelnen Items untersucht werden kann. Um festzustellen, ob Testaufgaben das theoretische Konstrukt abbilden, sind Informationen über die Antwortprozesse wichtig. Diese geben Aufschluss darüber, ob die kognitiven Prozesse zur Lösung der Items auf den vorab definierten Inhalt abzielen oder ob auch nicht intendierte irrelevante Prozesse zur Lösung führen können (vgl. Hartig et al. 2012, 161). Die von den Entwicklern intendierte fachliche Zielsetzung der Testaufgabe sollte auch von den Schülerinnen und Schülern in sinngleicher Weise verstanden werden (vgl. Kulgemeyer 2010, 88). Die konzipierten Items sollten sprachlich verständlich formuliert sein, weil semantische Missverständnisse als nicht intendierte Störvariablen auf die Testergebnisse einwirken. Die durch Sprache vermittelte inhaltliche Nachvollziehbarkeit der Items kann als Voraussetzung gesehen werden, eine Aufgabe richtig lösen zu können. Wenn beispielsweise Formulierungen, Fremdwörter oder einfach nur die Satzstellung bewirken, dass die Schülerinnen und Schüler den Inhalt der Aufgabe nicht verstehen, dann sollten neue Begrifflichkeiten gewählt bzw. Satzstrukturen geändert werden. Ausgenommen sind davon Fachbegriffe, die für eine richtige Lösung der Aufgabe relevant sind.

Bei gebundenen im Vergleich zu offenen Aufgabentypen besteht zudem die Herausforderung, dass die Antwortalternativen bereits vorgegeben sind und die Schülerinnen und Schüler ihre Lösung nur zwischen den bestehenden Antworten auswählen können. Sie sind in ihrem Antwortverhalten nicht frei (vgl. Jonkisz et al. 2012, 43), sondern müssen sich auf der Basis der gegebenen Möglichkeiten für eine Lösung entscheiden. Dabei decken die Antwortmöglichkeiten vielleicht nicht alle Reaktionsmöglichkeiten ab, beeinflussen sich gegenseitig oder entsprechen nicht der natürlichen Antwort des Befragten, was sich negativ auf die Validität auswirken kann (vgl. Rost 2004, 61). Um die Nichtwissenden von den Wissenden

unterscheiden zu können, sollten die Distraktoren von unwissenden Schülerinnen und Schülern auch tatsächlich als mögliche richtige Antworten angesehen werden. Aufgrund dessen erscheint es gerade bei gebundenen Antwortformaten bedeutsam, die Testaufgaben auf der Ebene der Items hinsichtlich ihrer Validität zu untersuchen.

Eine Option, einen Einblick in den Lösungsprozess erhalten und die Testitems hinsichtlich ihrer sprachlichen Angemessenheit prüfen zu können, ist die Methode „Lautes Denken“ (nach Ericsson/Simon 1984). Durch diese qualitative Methode können Gedanken und Strategien erhoben werden, die die Schülerinnen und Schüler nutzen, um ihre Antwort auszuwählen. Außerdem kann die inhaltliche Nachvollziehbarkeit der Item-Formulierungen bzw. die mögliche (In-)Kongruenz zwischen den Intentionen der Testkonstrukteure und den Vorstellungen der Testpersonen untersucht werden. Im vorliegenden Beitrag wird eine Studie vorgestellt, in der diese qualitative Methode eingesetzt wurde, um den Bearbeitungsprozess der Schülerinnen und Schüler sichtbar zu machen und die Verständlichkeit der Items zu prüfen. Wenn Erkenntnisse über den Antwortprozess vorliegen, dann kann analysiert werden, wie die Schülerinnen und Schüler zu ihrer Antwort gelangen und ob sie die Items in der vom Testkonstrukteur intendierten Weise verstehen, Verständnisprobleme vorliegen und schülernahe Vorstellungen in den Antwortmöglichkeiten enthalten sind. Durch die gewonnenen Informationen und anschließende Überarbeitung dieser, kann sich die Validität der Items erhöhen.

Die Grundlage für die vorliegende Untersuchung bilden Testitems, die zur Erfassung ökonomischen Fachwissens der Schuljahrgänge acht und neun an Haupt-, Real- und Oberschulen des Sekundarbereichs I entwickelt wurden (vgl. Loerwald/Schnell 2014). Diese Items wurden von Fachdidaktikerinnen und Fachdidaktikern erstellt, von Lehrpersonen bewertet und auf der Itemebene mithilfe der Methode „Lautes Denken“ durch einzelne Schülerinnen und Schüler untersucht. Die modifizierten Items wurden anschließend mit verschiedenen Stichproben getestet. Im Folgenden werden das „Lautes Denken“ als Forschungsmethode beschrieben, die Durchführung und die Ergebnisse zur Prüfung der Validität der entwickelten Testaufgaben dargelegt und Überlegungen zur Konzeption von Testaufgaben in der ökonomischen Bildung formuliert.

2 „Lautes Denken“ als Forschungsmethode

Unter der Methode „Lautes Denken“ wird „[...] das gleichzeitige laute Aussprechen von Gedanken bei der Bearbeitung einer Aufgabe [...]“ (Knoblich/Öllinger 2006, 692) verstanden.

Der Ursprung der Methode „Lautes Denken“ liegt laut Weidle und Wagner (1982, 81) in der Denkpsychologie von Bühler (1907), der diese Methode entwickelte, um die kognitiven Prozesse des Problemlösens zu entdecken. Erst im Rahmen der sogenannten „kognitiven Wende“ Mitte des letzten Jahrhunderts rückten grundsätzliche Fragen über den menschlichen Verstand, das Denken und die Verarbeitungsprozesse wieder in den Fokus von Lehr-Lern-Prozessen (vgl. Kucan/Beck 1997, 271).

Das „Laute Denken“ wird seither in verschiedenen Kontexten und besonders in der pädagogisch-psychologischen und fachdidaktischen Lehr-Lern-Forschung zur Analyse von Denk-, Lern- und Problemlöseprozessen verwendet (vgl. Sandmann 2014, 181). Im Vergleich zu anderen qualitativen Formen wird bei dieser Methode nicht nach der Meinung oder Einstellung der Person zu einem Sachverhalt gefragt, sondern es werden die kognitiven Prozesse, mentalen Operationen und Strategien erhoben, die während der Bearbeitung einer Aufgabe ablaufen (vgl. Weidle/Wagner 1982, 82; Knoblich/Öllinger 2006, 692). In der Kompetenzmodellierung und Testentwicklung bietet die Methode „Lautes Denken“ die Möglichkeit zur Überprüfung der Qualität von Testaufgaben, indem die Verständlichkeit der einzelnen Items getestet, Verständnisprobleme aufgedeckt sowie die Aufgabengüte und Item-Validierung geprüft werden (vgl. Sandmann 2014, 182). Das „Laute Denken“ gilt als eine der wenigen Methoden, die die informativen kognitiven Vorgänge erfasst, die während einer Handlung auftreten (vgl. Weidle/Wagner 1982, 83), und stellt eine wichtige Methode für die Erhebung von Denkprozessen dar (vgl. Knoblich/Öllinger 2006, 695).

In der Auseinandersetzung mit der Methode „Lautes Denken“ werden immer wieder zwei zentrale Kritikpunkte angeführt: (a) Die kognitiven Prozesse bei der Bearbeitung von Aufgaben seien so komplex, dass sie nicht durch simultanes Verbalisieren hinreichend vollständig erfasst werden können und (b) das simultane Verbalisieren beeinflusse die Bearbeitung der Items und führe zu einer Veränderung kognitiver Leistungen.

Zu (a): Weidle und Wagner (1982, 84), Yom et al. (2009, 639) geben die Kritik wieder, dass es eine Diskrepanz zwischen der gedanklichen Verarbeitung und dem gleichzeitigen Prozess der Verbalisierung gebe, weil das Entstehen von Gedanken in einem schnelleren Prozess verlaufe als die Umsetzung der Gedanken in Worte. Daher sei es gar nicht möglich, dass eine Schülerin/ein Schüler alle Gedanken, die ihr/ihm bei der Bearbeitung der Aufgabe durch den Kopf gehen, auch laut artikuliere. Das heißt, einige Gedanken, die ggf. zur Bearbeitung der Aufgabe beitragen, würden eventuell nicht geäußert. Dieser Kritikpunkt stellt die Methode an sich aber nicht infrage, sondern zeigt lediglich ihre Grenzen auf. Beim „Lauten

Denken“ kann zwar nicht sichergestellt werden, dass alle Gedanken einer Testperson bei der Bearbeitung einer Aufgabe auch laut geäußert werden, aber es können relevante Vorstellungen erhoben werden, die Hinweise für die sprachliche und inhaltliche Verbesserung der Testitems liefern. Die Übereinstimmung der verbalisierten Daten mit den tatsächlichen mentalen Prozessen und somit auch die Validität der Daten ist dann am höchsten, wenn eine simultane Verbalisierung der Gedanken bei der Bearbeitung der Aufgaben gefordert wird (vgl. Sandmann 2014, 188; Buber 2009, 562).

Zu (b): Weiterhin wird an der Methode kritisiert, dass das „Laute Denken“ den Problem- und Lösungsprozess beeinflusse und kognitive Leistungen verändere (siehe u. a. Funke/Spering 2006, 676). Als Hintergrund dieses Kritikpunktes kann u. a. das Prozessmodell von Ericsson und Simon (1980) gesehen werden, welches postuliert, dass verbale Daten direkt laut geäußert werden können, während nonverbale Daten (beispielsweise Bilder) zusätzlich in einem weiteren Schritt erst in Sprache umgewandelt werden müssen. Dieser zusätzliche Rekodierungsschritt würde zu einer verlangsamten Bearbeitung der Aufgaben und folglich zu einer Interferenz der Versprachlichung mit dem Problemlöseprozess führen (vgl. Funke/Spering 2006, 676). Mit der Kritik der kognitiven Veränderung durch das „Laute Denken“ beim Problemlöseprozess hat sich die empirische Forschung auseinandergesetzt. Studien wie die von Schooler et al. (1993) und Franzen/Merz (1998) konnten eine Veränderung der kognitiven Leistung durch das „Laute Denken“ belegen. Deffner (1989) konnte durch eine Studie zur Prüfung des Prozessmodells von Ericsson und Simon (1980) die Möglichkeit ausschließen, dass sich die Aufgabenbearbeitung zwischen Personen, die ihre Gedanken laut äußern, und denjenigen, die die Aufgaben still bearbeitet haben, qualitativ unterscheiden. Bise (2008) bestätigte die Ergebnisse von Ericsson/Simon (1980) und Deffner (1989) durch ihre Untersuchung der dialogischen Strukturen im inneren Sprechen beim Problemlösen. Die Bearbeitungsdauer bei bildhaften Items und Aufgaben mit höherem Schwierigkeitsgrad verlängert sich durch das „Laute Denken“, aber der Lösungsweg an sich verändert sich nicht (vgl. Bise 2008, 216f.). Flaherty (1975) konnte in einem entsprechenden Experiment ebenfalls keinen signifikanten Unterschied hinsichtlich des kognitiven Prozesses zwischen der Experimental- und Kontrollgruppe feststellen und Biggs et al. (1993) bestätigten diese Ergebnisse, da auch sie keinen Einfluss des „Lauten Denkens“ auf die kognitive Leistung identifizieren konnten. Zusammenfassend kann festgehalten werden, dass die gleichzeitige Verbalisierung der Gedanken während der Bearbeitung zu keiner Veränderung des kognitiven Vorgangs führt und

die Ergebnisse der Methode „Lautes Denken“ als valide und nicht-reaktiv bezeichnet werden können (vgl. Knoblich/Öllinger 2006, 693).

Wie bei vielen qualitativen und quantitativen Befragungsmethoden wird auch bei dieser explorativen Methode kritisiert, dass während der Durchführung die Kommunikationssituation unnatürlich sei (vgl. Leuders et al. 2011, 216). Die Person kann sich beobachtet, bewertet und getestet fühlen (vgl. Yom et al. 2009, 638). Zudem ist es möglich, dass die Jugendlichen versuchen, sich so zu verhalten und auszudrücken, dass sie die Erwartungen der Forscherin/des Forschers erfüllen (vgl. Weidle/Wagner 1982, 85). Die hier zuletzt beschriebenen Grenzen der Methode können durch einen Planungs- und Durchführungsablauf, der diese Kritikpunkte berücksichtigt [wie beispielsweise die verbale Darstellung der Testaufgaben (vgl. Silberer 2005, 264), die Schaffung einer störfreien Situation (vgl. Konrad 2010, 487) und das Einplanen einer Aufwärmphase (vgl. Ericsson/Simon 1998, 181)] minimiert werden. Trotz dieser kritischen Anmerkungen ist diese Methode eine der wenigen Methoden, die mentale Prozesse während einer Handlung erheben (vgl. Sandmann 2014, 188) und valide Ergebnisse ermitteln kann.

3 Zum Einsatz der Methode „Lautes Denken“ im Rahmen einer Studie zur Erfassung ökonomischen Fachwissens von Schülerinnen und Schülern

In den naturwissenschaftlichen Fachdidaktiken ist das „Laute Denken“ bereits eine gängige Methode zur Überprüfung der Qualität von Testaufgaben. Die Methode „Lautes Denken“ wird sowohl bei gebundenen Aufgabentypen (beispielsweise Ordered Multiple Choice, Forced-Choice-Aufgaben) als auch bei offenen Formaten zur Erfassung des Fachwissens, der Erkenntnis- oder der Bewertungskompetenz eingesetzt (siehe u. a. Hadenfeldt et al. 2014; Krell et al. 2012; Gausmann et al. 2010; Völzke 2012; Bernholt et al. 2012; Kulgemeyer 2010; Prenzel et al. 2007).

In der ökonomischen Bildung gibt es mittlerweile mehrere Studien, die – ähnlich wie die vorliegende es skizziert – mithilfe von standardisierten Tests den ökonomischen Kenntnisstand von Jugendlichen erheben (siehe u. a. Greimel-Fuhrmann et al. 2016; Brandlmaier et al. 2006; Wobker et al. 2012; Würth/Klein 2001; Beck/Krumm 1998) und seit einigen Jahren werden Kompetenzmodelle entwickelt, die mithilfe von Testverfahren operationalisiert und empirisch erprobt werden (vgl. u. a. Schumann et al. 2011; Hentrich/Seeber 2016; Macha 2015). Die Methode „Lautes Denken“ wurde – soweit bekannt – in den genannten Studien

aber noch nicht zur Untersuchung der Validität von Testaufgaben im Sekundarbereich I genutzt. Sie wurde bis dato lediglich – mit Ausnahme der in diesem Beitrag beschriebenen Studie – in einem Projekt zur Erfassung von ökonomischen Präkonzepten von Grundschulkindern eingesetzt (Arndt/Kopp 2016).

3.1 Theoretischer Rahmen und Fragestellungen

Die in der hier skizzierten Studie erstellten und empirisch erprobten Items sind Gegenstand von Testinstrumenten zur Erfassung des ökonomischen Fachwissens von Jugendlichen für den achten und neunten Jahrgang im Fach Wirtschaft der Haupt-, Real- und Oberschulen in Niedersachsen. Diese Testinstrumente sollen nach entsprechender Prüfung Lehrpersonen als Diagnoseinstrument im Fach Wirtschaft in Niedersachsen zur Verfügung gestellt werden. Den theoretischen Rahmen für die entwickelten Testaufgaben in dieser Untersuchung bilden daher die niedersächsischen Kerncurricula der Haupt-, Real- und Oberschule, die in gleicher Weise aufgebaut und nahezu identische Kompetenzen aufweisen. Dementsprechend sind die Testaufgaben inhaltlich in vier Themenfeldern verortet: 1) Verbraucherinnen und Verbraucher sowie Erwerbstätige im Wirtschaftsgeschehen, 2) Ökonomisches und soziales Handeln in Unternehmen, 3) Die Aufgaben des Staates im Wirtschaftsprozess, 4) Ökonomisches Handeln, regional, national und international (vgl. Niedersächsisches Kultusministerium 2009a; Niedersächsisches Kultusministerium 2009b; Niedersächsisches Kultusministerium 2013).

Die Testaufgaben fokussieren den in den Kerncurricula ausgewiesenen Kompetenzbereich „Fachwissen“, der in allen drei Lehrplänen identisch formuliert ist. Unter Fachwissen wird die Fähigkeit verstanden „[...] über strukturiertes ökonomisches Wissen zu verfügen und damit gesellschaftliche Strukturen und Prozesse zu verstehen“ (Niedersächsisches Kultusministerium 2009a, 14). Die Jugendlichen sollen über ökonomische Kenntnisse verfügen, die sie befähigen „[...] ihre Interessen und Bedürfnisse, ihre Handlungsspielräume und Entfaltungsmöglichkeiten im Zusammenhang mit real vorhandenen ökonomischen Strukturen und Mechanismen zu sehen“ (ebd.).

Die Items wurden als Single-Choice-Aufgaben (vierstufiges Aufgabenformat mit einem Attraktor) unter Berücksichtigung testtheoretischer Kriterien entwickelt (siehe Loerwald/Schnell 2014). In einem zweiten Schritt wurden alle Testaufgaben von Lehrpersonen u. a. hinsichtlich der curricularen und inhaltsbezogenen Validität, der Schwierigkeit und sprachlichen Verständlichkeit bewertet. Basierend auf diesen Beurteilungen wurden die Items modifiziert,

welche dann die Grundlage für den Einsatz der Methode „Lautes Denken“ bildeten. Durch die Überprüfung der Testaufgaben mithilfe der explorativen Methode sollten folgende zwei Fragestellungen untersucht werden:

- a) Welche Strategien nutzen die Schülerinnen und Schüler zur Auswahl ihrer Lösung?
- b) Inwiefern sind die entwickelten Aufgaben für die Schülerinnen und Schüler des achten und neunten Jahrgangs sprachlich verständlich formuliert?

3.2 Methodisches Vorgehen und Sampling

In der vorliegenden Untersuchung wurde die simultane Verbalisierung der Gedanken (per-aktionale Variante) nach Ericsson und Simon (1984) verwendet. Die Schülerinnen und Schüler haben ihre Gedanken während der Auseinandersetzung mit der Aufgabe verbalisiert, wodurch es möglich war, den Gedankenprozessen zu folgen und mögliche Verständnisprobleme sowie Lösungsstrategien der Aufgaben aufdecken zu können. Für die Nachbesprechung einzelner Testitems wurde zusätzlich auf die retrospektive Variante des „Lauten Denkens“ zurückgegriffen, um ggf. Gedanken und Gedankenprozesse klären zu können (siehe Konrad 2010, 477).

In den Jahren 2012 und 2014 wurden für den Jahrgang acht drei Schülerinnen und drei Schüler und für den Jahrgang neun drei Schüler (eine Schülerin und zwei Schüler) aus dem Sekundarbereich I (Oberschule, Integrierte Gesamtschule) für die Überprüfung der Testaufgaben ausgewählt. Bei der Auswahl des Samplings wurde darauf geachtet, dass die Schülerinnen und Schüler in der Lage sind, ihre Gedanken laut zu äußern, weshalb eher extrovertierte Schülerinnen und Schüler in die Untersuchung aufgenommen wurden. Zudem wurden Jugendliche mit Interesse an Wirtschaft, aber mit unterschiedlichen fachlichen Leistungsniveaus ausgewählt. Den Schülerinnen und Schülern wurde ein Testheft mit insgesamt 36 Aufgaben (2012) bzw. 27 Aufgaben (2014) für den achten Jahrgang und 40 Items für den neunten Jahrgang (2014) vorgelegt.

3.3 Datenerhebung

Die Durchführung erfolgte durch wissenschaftliche und studentische Mitarbeiterinnen und Mitarbeiter an mehreren Vormittagen an den jeweiligen Schulen. Die Überprüfung der Aufgaben wurde von einer Forscherin/einem Forscher mit jeweils nur einem Jugendlichen (Einzeltestung) in einem dafür gebuchten (störfreien) Raum durchgeführt. Vor den Durchführungen erfolgte eine Schulung der Forschergruppe, in der der Leitfaden besprochen wurde, um eine standardisierte Durchführung der Methode „Lautes Denken“ sicherstellen zu können.

Zu Beginn der Durchführung der Methode „Lautes Denken“ wurde eine angenehme und freundliche Atmosphäre geschaffen und Zeit für ein Kennenlernen eingeräumt, um diese eher unnatürliche Situation etwas zu entschärfen. Um der Schülerin/dem Schüler nicht das Gefühl zu geben, dass sie/er als Person beobachtet und getestet wird, wurde zu Beginn das Ziel und die Methode „Lautes Denken“ erläutert und betont, dass der Fokus nicht auf der Bewertung der einzelnen Schülerin/des einzelnen Schülers liegt, sondern auf der Überprüfung der Testaufgaben. Auf zu spezifische Informationen über den Zweck der Durchführung wurde verzichtet, um Spekulationen über das, „was der Forscher hören möchte“, zu vermeiden (vgl. Englund Dimitrova 2005, 72). Die Anweisungen, wie die Methode „Lautes Denken“ umgesetzt werden soll, wurden zu Beginn eindeutig und klar formuliert (siehe Ericsson/Simon 1984, 80) sowie der Schülerin/dem Schüler schriftlich vorgelegt und gemeinsam besprochen (siehe Göpferich 2007, 2). Die Instruktion zum „Lauten Denken“ enthielt die Aufforderung, die Gedanken direkt zu äußern und diese nicht zu selektieren oder zu filtern, um reliable und valide Daten erhalten zu können (siehe dazu Knoblich/Öllinger 2006, 695). Bevor die entwickelten Testaufgaben mithilfe der Methode „Lautes Denken“ von den Schülerinnen und Schülern bearbeitet wurden, erfolgte eine Trainingsphase, in der die Verbalisierung der Gedanken anhand von mathematischen Beispielaufgaben eingeübt wurde (siehe u. a. Sandmann 2014; Ericsson und Simon 1984). Hierfür wurden offene und gebundene Items eingesetzt, beispielsweise eine Multiplikationsaufgabe wie von Ericsson und Simon (1998, 81) vorgeschlagen. Die Schülerin/der Schüler erhielt dadurch die Möglichkeit, die Methode „Lautes Denken“ zu üben und sich mit der eher unnatürlichen Kommunikationssituation auseinanderzusetzen zu können.

Während der gesamten Testdurchführung befand sich die jeweilige Person außerhalb des Sichtfeldes der Schülerin/des Schülers. Wenn die Schülerin/der Schüler an das Verbalisieren der Gedanken erinnert werden musste, wurde „Lautes Denken nicht vergessen“ in Form einer Aussage und nicht als Frage formuliert (siehe Ericsson/Simon 1984, 80). Diese Erinnerung erfolgte nach einer Pause von etwa zwölf Sekunden. Die Schülerinnen und Schüler erhielten nach der Bearbeitung des Tests die Möglichkeit, eine Reflexion zu dieser für sie eher ungewöhnlichen Situation zu geben (siehe Frommann 2005, 4). Außerdem wurden offene Fragen zu einzelnen Äußerungen und Testaufgaben im Anschluss an die Testdurchführung mit der retrospektiven Variante der Methode „Lautes Denken“ erfasst. Die durchschnittliche Bearbeitungszeit und das nachträgliche Besprechen der noch offenen Fragen zu

den Testaufgaben lagen bei dem achten Jahrgang bei 55 bzw. 34 Minuten und im neunten Jahrgang bei 46 Minuten.

3.4 Auswertung und exemplarische Ergebnisse

Die Durchführung der Methode „Lautes Denken“ wurde aufgezeichnet und zu unstrukturierter, simultanen Denke-Laut-Protokollen mit dem Programm „f4“ transkribiert. Hierbei wurde ein einfaches Transkriptionssystem genutzt und die Bedeutung „Wie-etwas-gesagt-wurde“ (Dresing/Pehl 2011, 25) vorerst nicht ausgewertet, da es für die Bewertung und Prüfung der Testaufgaben in dieser Phase nicht entscheidend war. Die Äußerungen wurden wortwörtlich transkribiert und die Länge der Pausen zwischen den Antworten mit Klammern gekennzeichnet.

Laut Mackensen-Friedrichs (2004) stellt für die Schülerinnen und Schüler der Klasse neun das laute Äußern von Gedanken keine Schwierigkeit dar (vgl. ebd., 82). Allerdings zeigte sich bei der vorliegenden Untersuchung, dass insbesondere die Schülerinnen und Schüler (beider Jahrgänge), die über das notwendige Wissen zur korrekten Beantwortung der Items verfügten, Schwierigkeiten hatten, ihre Gedanken simultan zu äußern, da für sie die Lösung eindeutig war und keiner Ausführung bedurfte. Durch das Erinnern an das „Laute Denken“ während der Bearbeitung und das retrospektive Besprechen der Aufgaben wurde versucht, dieses Problem zu entschärfen. Ein Schüler zeigte starke Schwierigkeiten mit der Methode „Lautes Denken“ und äußerte nur die seiner Meinung nach richtige Antwort der Testaufgaben, ohne die jeweiligen Gedanken während der Aufgabenbearbeitung zu verbalisieren. Auch durch die Aufforderung „Lautes Denken nicht vergessen“ während der Bearbeitung und das nachträgliche „Laute Denken“ konnten keine Informationen über die Lösungsprozesse gewonnen werden. Das Transkript musste von der Auswertung ausgeschlossen werden.

Als Auswertungsmethode wurde die qualitative Inhaltsanalyse nach Mayring (2010) gewählt, die es ermöglicht, Material zu untersuchen, dieses auf das Wesentliche zu reduzieren und nur ausgewählte Textteile zu berücksichtigen (vgl. Mayring 2010, 66). Mithilfe der Variante der zusammenfassenden Inhaltsanalyse konnte das erhobene Datenmaterial systematisch, regel- und theoriegeleitet ausgewertet werden, was wiederum Rückschlüsse auf den ermittelten Gegenstand erlaubte (vgl. Mayring 2013, 472). Bei der hier gewählten induktiven Vorgehensweise wurden vorab basierend auf den Fragestellungen Selektionskriterien formuliert

(vgl. ebd., 84; Kirchner 2016, 194). Drei Selektionskriterien wurden aufgrund der Fragestellungen der Untersuchung für die induktive Kategorienbildung definiert:

- Strategien zur Auswahl der Lösung
- Plausibilität der Distraktoren
- Verständlichkeit der Testaufgabe

Eine Aussage dazu, wie die richtige oder scheinbar richtige Antwort ausgewählt worden war, wurde dem Selektionskriterium „Strategien zur Auswahl der Lösung“ zugeordnet. Aussagen der Schülerinnen und Schüler, die Informationen zur Qualität der falschen Antworten im Test lieferten, wurden dem Kriterium „Plausibilität der Distraktoren“ zugeordnet. Aussagen der Jugendlichen, ob die Items sprachlich nachvollziehbar waren, wurden dem Kriterium „Verständlichkeit“ zugewiesen. Die Ausprägungen dieser Selektionskriterien wurden dann induktiv aus dem aufgezeichneten Material gebildet. In dieser Studie wurden alle expliziten Äußerungen der Schülerinnen und Schüler zu den beiden Forschungsfragen bestimmt und als Analyseeinheiten wurden ganze Sätze definiert. Das Datenmaterial wurde nach den einzelnen Selektionskriterien von zwei Wissenschaftlerinnen unabhängig voneinander analysiert, konkrete Textpassagen notiert, paraphrasiert, reduziert und die entwickelten Kategorien anschließend diskutiert. Die Anzahl der übereinstimmenden kodierten Schüleraussagen wurde nach der Formel von Holsti (1969) berechnet und weist einen Wert von 0.84 auf. Ausgehend von den Selektionskriterien „Strategien zur Auswahl der Lösung“, „Plausibilität der Distraktoren“ und „Verständlichkeit der Testaufgabe“ wurden die Kategorien *Fachlichkeit*, *sprachliche Analogiebildung*, *Ausschlussprinzip* und *Verständlichkeit* gebildet. Die Ergebnisse der Analyse werden im Folgenden ausgeführt.

3.4.1 Strategien der Schülerinnen und Schüler zur Auswahl ihrer Lösungen

Durch die Analyse konnte aufgezeigt werden, dass die Schülerinnen und Schüler unterschiedliche Strategien anwendeten, um ihre Antwort auszuwählen. Schülerinnen und Schüler beantworteten Items inhaltlich korrekt aufgrund ihres vorhandenen ökonomischen Fachwissens (Kategorie: *Fachlichkeit*). Bei fehlendem Wissen wurden andere Strategien (Kategorien: *sprachliche Analogiebildung*, *Ausschlussverfahren*) genutzt, um zu einer Antwort zu gelangen.

Die meisten Testaufgaben wurden von den Schülerinnen und Schülern fachlich korrekt begründet und mithilfe ihres ökonomischen Wissens richtig gelöst (Kategorie: *Fachlichkeit*). Im

Folgendes wird eine Testaufgabe des neunten Jahrgangs und die laut geäußerten Gedanken eines Schülers exemplarisch für diese Kategorie angeführt:

Beispielitem: Beim winterlichen Einkauf auf dem Wochenmarkt bleiben Peter und sein Vater an einem Obst- und Gemüsestand stehen und werfen einen Blick auf die angebotenen Erdbeeren. Eine Schale Erdbeeren (500 g) wird für 4,50 Euro angeboten und liegt somit 2,- Euro über dem Preis im Juni. Kreuze die Antwort an, die beschreibt, wie die Marktpreise der Erdbeeren zustande gekommen sind. Die unterschiedlichen Marktpreise für Erdbeeren kommen zustande, da ...

- a) *im Winter die Preise für alle Güter höher sind.*
- b) *die Anbieter im Winter einen größeren Gewinn erzielen können.*
- c) *im Sommer die angebotenen Mengen an Erdbeeren größer sind.*
- d) *die Nachfrage nach Erdbeeren im Winter größer ist.*

Auszug aus der Schülerantwort:

„Erdbeeren wachsen, denke ich mal, im Winter nicht so gut, deswegen (...) ist die/ (..) sind die angebotenen Erdbeeren dann mehr und dann können die die nicht alle verkaufen und müssen mit dem Preis runtergehen. Wenn dann im Winter weniger sind, dann wird das wieder mehr (..) was die kosten.“ (Schüler der 9. Klasse im Jahr 2014)

Der Schüler kreuzte die fachlich richtige Antwort an und das Item wurde sprachlich verstanden. In diesem Beispiel führten bei dem Schüler keine unbeabsichtigten Prozesse zur Lösung. Bei fehlendem Wissen hingegen waren zwei zentrale Strategien der Schülerinnen und Schüler aus dem Sampling erkennbar, die den Kategorien (a) *sprachliche Analogiebildung* und (b) *Ausschlussprinzip* zugeordnet wurden.

Zu Kategorie (a) *sprachliche Analogiebildung*: Die Schülerinnen und Schüler haben Begriffe und Wörter, die sie aus ihrem Alltag kannten, auf die ökonomische Situation der Testaufgaben übertragen. Hierbei wurde die jeweilige Bedeutung des Wortes aus der Alltagssprache mit dem ökonomischen Fachbegriff gleichgesetzt, was teilweise zu einer fachlich nicht korrekten Interpretation führte. Beispielsweise wurde bei einer Aufgabe zu fixen und variablen Kosten die alltägliche Bedeutung des Begriffs analog zu der ökonomischen Bedeutung verwendet und die Lösung unter Berücksichtigung der lebensweltlichen Erfahrung ausgewählt, was im Folgenden näher beschrieben wird.

Bei dem Pretest zur Methode „Lautes Denken“ zeigte sich, dass den Jugendlichen die Begriffe „fixe“ und „variable“ aus dem Unterricht nicht bekannt waren. Nach Auskunft der Jugendlichen bei der Nachbesprechung hatten sie im Unterricht stattdessen die Wörter „feste“

und „veränderliche“ Kosten verwendet. Da mit dieser Testaufgabe nicht die Definition der Fachbegriffe „fixe“ und „variable“ Kosten abgeprüft werden sollte, sondern ob die Schülerinnen und Schüler Kosten nach diesen beiden Arten unterscheiden konnten, wurden in der Hauptuntersuchung die aus dem Unterricht bekannten Begriffe „feste“ und „veränderliche“ als Hilfestellung in Klammern mitaufgeführt. Diese Single-Choice-Aufgabe mit vier Antwortmöglichkeiten enthielt Beispiele, die eindeutig nach variablen bzw. fixen Kosten eingeordnet werden konnten.

Bei der Hauptuntersuchung mithilfe der Methode „Lautes Denken“ zeigten sich allerdings durch diese Ergänzungen Schwierigkeiten. Eine Schülerin verwendete die Fachbegriffe der ökonomischen Bildung analog zu der ihr alltagssprachlich bekannten Bedeutung. Bei der Bearbeitung der Testaufgabe zu „variablen“ (veränderlichen) Kosten äußerte die Schülerin, nachdem sie den Aufgaben- und Itemstamm sowie die gegebenen Antwortalternativen [a) Rundfunkgebühren; b) Versicherungen; c) Getränke; d) Beitrag im Sportverein] laut vorgelesen hatte, laut folgende Gedanken:

„[...] Also (---) Beiträge im Sportverein, das (.) kann man ja verändern, man kann ja (2s) auch z. B. aus dem Sportverein austreten (---).“ Zudem wurde unter Berücksichtigung der Erfahrungen aus der Lebenswelt weiter begründet: [...] aber die Versicherung ja z. B. braucht man, hat man eigentlich das Leben lang. Getränke braucht man auch, weil man ohne (.) Wasser z. B. auch nicht (--) überleben kann (2s). Beim Rundfunkgebühr weiß ich gar nicht. [...] Aber ich glaub, dass muss man einfach bezahlen. [...] Also sage ich jetzt Beitrag im Sportverein [2s] weil man da ja auch austreten kann.“ (Schülerin der 8. Klasse im Jahr 2012).

Die ökonomische Bedeutung der Begriffe „variable“ (veränderliche) und „fixe“ (feste) Kosten waren der Schülerin anscheinend nicht bekannt. Sie deutete den Begriff „veränderliche“ Kosten in ähnlicher Weise, wie sie diesen aus ihrer alltagssprachlichen Bedeutung kannte und bildete eine Analogie. Die Schülerin verstand „variable (veränderliche) Kosten“ eher als Kosten, die die einzelne Person für sich selbst ändern kann und überlegte, ob die als Antwortmöglichkeiten genannten Kosten für das eigene Leben einen entscheidenden Einfluss haben. Sie hatte sich unter Berücksichtigung ihres lebensweltlichen Wissens hergeleitet, dass Versicherungen ein fester Bestandteil des Lebens sind, Wasser zum Leben notwendig ist, Rundfunkgebühren zu zahlen sind und dass diese Kosten für das eigene Leben nicht entfallen (verändert werden) können bzw. sollten. Sie hatte sich aufgrund ihrer Vorstellungen für die für sie vermeintlich richtige Antwort entschieden, die unter Berücksichtigung der ökonomischen Bedeutung falsch ist.

Nach der Auswertung dieser Ergebnisse wurde erstens entschieden, die beiden Hilfwörter (feste und veränderliche) beizubehalten, da diese Begriffe eher im Unterricht verwendet werden und nicht das Abprüfen der Definition der Fachbegriffe „fixe“ und „variable“ im Fokus stand, sondern der Transfer auf vorgegebene Beispiele. Aus den beiden bestehenden Testaufgaben (jeweils eine Aufgabe zum Abprüfen des Fachwissens zu „fixen“ und eine zu „variablen“ Kosten) wurde nur noch eine Testaufgabe zu „fixen“ Kosten generiert, da aufgrund der Relevanz des Inhalts ein Item als ausreichend angesehen wurde und damit keine Abhängigkeit zwischen den Items entstehen konnte. Zweitens wurde als Konsequenz für die Testaufgabe „fixe Kosten“ überlegt, welche Antwortmöglichkeiten dieser Strategie der Analogiebildung folgen und als Antwortalternativen genutzt werden könnten, um sicherzustellen, dass zum einen plausible Distraktoren formuliert werden, um den ratenden von dem wissenden Jugendlichen unterscheiden zu können, und zum anderen die fachlich korrekte Lösung nicht auch gleichzeitig die richtige Antwort ist, wenn Schülerinnen und Schüler sie durch die Bildung von Analogien herleiten. Als richtige Antwort wurde der fixe Kostenpunkt a) Miete gewählt und als falsche Antworten wurden die Beispiele b) Lebensmittel, c) Fahrradreparatur und d) Kleidung bestimmt. Diese Distraktoren sollten von Schülerinnen und Schülern, die sich bei fehlendem Wissen den Attraktor über die Analogiebildung des Wortes und durch Erfahrungen aus der Lebenswelt erschließen, als mögliche richtige Antwort angesehen werden, da Kleidung und Lebensmittel zum Leben (in einer Gesellschaft) entscheidend sind und abhängig von den Lebensumständen auch die Reparatur des Fahrrads. Die richtige Antwort (Miete) könnte als falsch angesehen werden, da sich diese Situation, beispielsweise durch den Kauf einer Immobilie, ändern kann. Die Testaufgabe wurde anschließend empirisch mit verschiedenen Stichproben getestet und es konnte eine gute Qualität der einzelnen Antwortmöglichkeiten belegt werden.

Die Strategie, sich bei fehlendem Wissen durch sprachliche Analogiebildung die ökonomischen Begriffe unter Berücksichtigung der Lebenswelt zu erschließen, zeigte sich auch bei ökonomischen Komposita. In diesen Fällen zerlegten die Jugendlichen vorab die ihnen unbekanntes ökonomischen Fachbegriffe in einzelne Teilwörter. Diese setzten die Schülerinnen und Schüler dann analog zu ihnen bekannten Begriffen und erschlossen sich so die inhaltliche Bedeutung. Beispielsweise wurde der Begriff „Wirtschaftsordnung“ in zwei Wörter unterteilt. Der Begriff „Wirtschaft“ wurde dann von einer Schülerin (9. Jahrgang, 2014) mit „Unternehmen“ übersetzt und „Ordnung“ als die allgemeine Ordnung (Struktur) in einem Unternehmen verstanden. Die Schülerin prüfte bei der Bearbeitung des Items, ob die Antwort-

alternativen Beschreibungen enthielten, die sich auf Ordnungsstrukturen in Unternehmen bezogen. Da durch die Überprüfung der Items mithilfe der Methode „Lauter Denken“ identifiziert werden konnte, dass einige Schülerinnen und Schüler die Begriffe „Wirtschaft“ und „Unternehmen“ synonym verwendet hatten, wurden im Anschluss an die Auswertung die Distraktoren dieses Items so verändert, dass sie nun Aussagen zu Ordnungen in Unternehmen enthielten, um natürliche Antworten von Jugendlichen anzubieten und das Raten der richtigen Antwort zu erschweren. Damit sollte erreicht werden, dass die Testaufgabe das Wissen über den Begriff Wirtschaftsordnung erfasst und nicht der Attraktor als einzelne plausible Antwort angesehen wurde und somit Jugendliche die Aufgabe richtig lösten, obwohl sie diesen Begriff nicht kannten.

Wie bereits beschrieben, sind die Schülerinnen und Schüler bei gebundenen Formaten in ihrer Antwort nicht frei, sondern aufgefordert, aus den bestehenden Alternativen eine Antwort auszuwählen. Durch das „Laute Denken“ der Schülerinnen und Schülern bei der Bearbeitung der Testaufgaben konnte z. T. entdeckt werden, dass bei einigen Antworten aus Schülersicht die vermeintlich richtige Antwort fehlte. Beispielsweise sollte mithilfe einer Testaufgabe geprüft werden, ob die Merkmale der Planwirtschaft (Jahrgang 9) bekannt waren. Eine Schülerin wählte ebenfalls bei der Bearbeitung dieser Aufgabe die Strategie der Wortzerlegung und trennte den Begriff Planwirtschaft in zwei Wörter und schlussfolgerte „[...] *da wird ja was geplant* [...]“ (Schülerin der 9. Klasse). Da keine der vorgegebenen Antwortmöglichkeiten zu diesem Zeitpunkt Verben wie planen, organisieren o. Ä. und keinen Bezug zu Unternehmen enthielt, führte dies bei der Schülerin zu Verunsicherung bei der Auswahl ihrer Antwort. Als Konsequenz dieses Ergebnisses wurden anschließend die Antwortalternativen modifiziert und Verben wie „planen“ eingepflegt. Diese Wörter wurden aufgenommen, um die Vorstellungen von Jugendlichen aufzugreifen, die durch sprachliche Analogiebildung sich die Antwort herleiten, um im Sinne der Schülerinnen und Schüler natürliche Antwortalternativen zu generieren.

Bei einem Item zur Erfassung der Definition von öffentlichen Unternehmen verbalisierte ein Schüler des achten Jahrgangs bei der Bearbeitung der Aufgabe, dass dies all die Unternehmen seien, die in der Öffentlichkeit bekannt sind, so wie Nike und Adidas. Bei der Konstruktion eines Items zu öffentlichen Unternehmen bestand bei den Aufgabenentwicklern nicht die Vorstellung, dass Schülerinnen und Schüler die Bezeichnung „öffentlich“ in dieser Weise verstehen könnten. Nach der Durchführung der Methode „Lauter Denken“ wurde ein

Distraktor formuliert, der diese Schülervorstellung beinhaltet, um die Aufgabenqualität zu erhöhen.

Zu Kategorie (b) *Ausschlussprinzip*: Bei einigen Aufgaben schlossen Schülerinnen und Schüler, die nicht über das fachliche Wissen zur Beantwortung der Aufgabe verfügten, die einzelnen Distraktoren aus. Exemplarisch kann dies an einer Aufgabe zur bargeldlosen Bezahlung (Jahrgang 9) gezeigt werden:

Beispielitem: Wiebke (19 Jahre) nutzt verschiedene Möglichkeiten des bargeldlosen Zahlungsverkehrs. Kreuze die richtige Aussage an.

- a) *Ein Dauerauftrag wird für das regelmäßige Bezahlen variierender Beträge genutzt.*
- b) *Eine Einzugsermächtigung erfolgt beim Bezahlen mit der EC-Karte.*
- c) *Eine Kreditkarte wird wie eine Geldkarte aufgeladen.*
- d) *Eine Überweisung ist eine einmalige Übertragung eines Geldbetrags.*

Die Schülerin argumentierte nach dem lauten Vorlesen der Aufgabe wie folgt:

„Also, eine Geldkarte, das ist ja praktisch so wie eine Taschengeldkarte und die wird ja jeden Monat aufgeladen und das ist ja bei einer/ (.) ja/eine Kreditkarte wird ja nicht (.) aufgeladen, sondern eine Kreditkarte ist womit du deine Kredite zahlst. Ein Dauerauftrag (.) wird nur mit (.) einem bestimmten (..) mit einer bestimmten Sache gezahlt, deswegen meine ich, dass es dann eine Einzugsermächtigung erfolgt, beim Bezahlen der EC-Karte.“ (Schülerin der 9. Klasse im Jahr 2014)

Die Schülerin entschied sich nach den genannten Begründungen, ohne Beachtung der vierten Antwortmöglichkeit, für die falsche Antwort b).

Nach der Durchführung erfolgte eine Nachbesprechung und es wurde bei dieser Aufgabe deutlich, dass der Begriff Kreditkarte nicht bekannt war, sondern die Schülerin aus ihrer Lebenswelt eine Prepaid-Kreditkarte kannte. Die Begriffe „Übertragung“ und „variieren“, die sie zuvor bei der Aufgabenbearbeitung kaum berücksichtigt bzw. ignoriert hatte, kannte sie nicht bzw. konnte sie nicht erklären. Die Schülerin hatte eine Antwort aufgrund der Analogiebildung und zwei andere Antwortalternativen aufgrund sprachlicher Verständnisprobleme ausgeschlossen und sich dann für die noch übrig gebliebene (falsche) Antwort entschieden. Nach der Auswertung wurden die Antwortalternativen dahingehend geändert, dass die Kreditkarte durch eine Prepaid-Kreditkarte ersetzt wurde, da die Kreditkarte nicht in ihrer Lebenswelt verankert war und die Jugendlichen eher die Prepaid-Kreditkarte zu kennen scheinen. Die Antworten wurden sprachlich vereinfacht, da die unbekannteren Begriffe für das Abprüfen irrelevant waren, und wie folgt geändert: a) Mit einer Prepaid-Kreditkarte kann man

nur online bezahlen; b) Eine Einzugsermächtigung wird durch das Bezahlen mit der EC-Karte gegeben; c) Mit einer Überweisung kann man einen Geldbetrag auf ein Konto einbezahlen; d) Eine Lastschrift ist ein Betrag, den der Käufer an den Verkäufer auf das Konto einbezahlt. Zudem war dieses Item sehr eingekleidet, weshalb der Stamm geändert wurde: „Es gibt verschiedene Möglichkeiten bargeldlos zu bezahlen. Kreuze die richtige Aussage an.“

Das Anwenden eines Ausschlussverfahrens bei gebundenen Formaten lässt sich nicht unterbinden, allerdings sollten durch gut formulierte und sprachlich verständliche Stämme und Antwortmöglichkeiten gewährleistet werden, dass Jugendliche die Aufgaben verstehen und die Antwortalternativen als natürlich ansehen, weshalb dieses Item in der beschriebenen Weise geändert wurde.

3.4.2 Verständlichkeit

Bei der Durchführung der Methode „Lautes Denken“ und insbesondere bei der anschließenden Besprechung der Testaufgaben zeigte sich, dass einige der Testaufgaben, trotz mehrfacher Prüfung von verschiedenen beteiligten Akteuren (u. a. Lehrpersonen), nicht zielgruppenadäquat formuliert waren (Kategorie *Verständlichkeit*). Es zeigten sich durch das „Laute Denken“ bei der Aufgabenbearbeitung Verständnisprobleme bei Satzkonstruktionen sowie bei einzelnen Verben und Substantiven. Bereits bei der Entwicklung der Testaufgaben wurde entsprechend der Zielgruppe u. a. versucht, kurze und einfach verständliche Sätze zu formulieren und doppelte Verneinungen zu vermeiden (vgl. Jonkisz et al. 2012, 64; Schmidt-Atzert/Amelang 2012, 90f.; Schnell et al. 2011, 328f.), um keine sprachlichen Hürden aufzubauen, sondern um das reine Fachwissen abprüfen zu können. Durch das laute Lesen und Äußern der kognitiven Prozesse während der Bearbeitung wurde erkennbar, dass einzelne Sätze noch zu lang und zu abstrakt formuliert waren und die Jugendlichen die inhaltliche Aussage nicht verstanden hatten. Aufgabenstämme wurden während der Aufgabenbearbeitung von Schülerinnen und Schülern mehrfach vorgelesen, einzelne, für die Lösung entscheidende Wörter überlesen, unverständliche Sätze ignoriert und es wurde verbalisiert, dass der beschriebene Inhalt oder die Aufgabenanweisung einzelner Aufgaben kognitiv nicht verstanden wurde. Es zeigte sich, dass einzelne Fachbegriffe wie beispielsweise „Inflationsrate“ und „Faktoren“ (Jahrgang 8), die als Distraktoren genutzt wurden, Jugendliche nicht bekannt waren. Darüber hinaus kannten einige der Befragten Verben wie beispielsweise „variieren“ und „objektiv“ nicht, die im Stamm oder als Antwortmöglichkeiten verwendet wurden. Diese Begriffe waren für das Lösen der jeweiligen Aufgabe nicht relevant. Schülerinnen

und Schüler, die die Testaufgaben aufgrund ihres bestehenden Wissens hätten richtig lösen können, aber diese sprachlich nicht verstanden haben, wurden u. U. dadurch benachteiligt. Diese Verständnisprobleme könnten sich negativ auf die Validität auswirken. Aufgrund dessen wurden nach der Durchführung der Methode „Lautes Denken“ diese Testaufgaben analysiert und anschließend die Länge der Aufgabenstämme gekürzt sowie die einzelnen Sätze knapper formuliert. Zudem wurden Wörter, die für das Abprüfen des jeweiligen Fachwissens irrelevant und den Jugendlichen unbekannt waren, durch andere Wörter ersetzt bzw. es wurde die Bedeutung der Begriffe umschrieben.

Es konnte festgestellt werden, dass einige Schülerinnen und Schüler sich bei Nichtwissen für die Antworten entschieden, die im Vergleich zu den anderen Antwortmöglichkeiten sehr konkret formuliert waren oder die einen Begriff enthielten, der auch im Aufgabenstamm enthalten war. Wurde bei einer Antwortmöglichkeit ein konkretes Beispiel formuliert, während hingegen bei allen anderen abstraktere und verallgemeinerte Aussagen beschrieben waren, tendierten die befragten Jugendlichen zu den konkreter formulierten Aussagen und gaben diese Antwort als Lösung an. Aufgrund dieser Erkenntnis wurden nach der Auswertung der Transkripte alle bereits erstellten Testaufgaben dahingehend geprüft, ob beispielsweise alle Antwortalternativen in ähnlicher Weise formuliert waren, die sich semantisch und formal ähnelten und ob sie auf einem annähernd gleichen Abstraktionsniveau lagen.

3.5 Zusammenfassung und Diskussion der Ergebnisse

In dieser Studie wurde die qualitative Methode „Lautes Denken“ eingesetzt, um die Validität auf der Ebene der Items zu untersuchen. Durch das simultane „Laute Denken“ der Schülerinnen und Schüler während der Bearbeitung der Items (in Einzeltestung) konnten verschiedene Strategien zur Auswahl der Lösung identifiziert und die Testaufgaben hinsichtlich ihrer Verständlichkeit und Kongruenz zwischen den Vorstellungen der Schülerin/der Schüler und der Testentwickler geprüft werden. Die hier beschriebenen Aufgaben stellen exemplarisch die Kategorien dar und geben einen Einblick, wie die entwickelten Testaufgaben von den Schülerinnen und Schülern gelöst wurden.

Die überwiegende Mehrheit der konzipierten Testaufgaben, die bereits von verschiedenen Lehrpersonen hinsichtlich der Verständlichkeit bewertet wurden, wurde von den Schülerinnen und Schülern der beiden Jahrgänge fachlich korrekt beantwortet und die Auswahl der Lösung fachlich korrekt begründet und sprachlich verstanden (Kategorien *Fachlichkeit* und

Verständlichkeit). Die Schülerinnen und Schüler begründeten und wählten aufgrund ihres vorhandenen ökonomischen Fachwissens die korrekte Antwort aus.

Durch das „Laute Denken“ konnte weiter identifiziert werden, dass Schülerinnen und Schüler bei fehlendem ökonomischen Fachwissen auch andere Strategien nutzten, um zu einer Antwort zu gelangen. Diese laut geäußerten Gedanken der Jugendlichen bei der Bearbeitung der Testaufgaben wurden nach den Kategorien *sprachliche Analogiebildung* und *Ausschlussverfahren* differenziert. Es zeigte sich, dass die Schülerinnen und Schüler die ökonomischen Fachbegriffe analog zu den Begriffen setzten, die sie bereits kannten und mithilfe von lebensweltlichen Erfahrungen versuchten, die richtige Antwort zu entdecken und teilweise per Ausschlussverfahren ihre Antwort auswählten. Einige Testaufgaben enthielten unbekannte Wörter und wurden sprachlich nicht verstanden, was sich durch die laut geäußerten Gedanken zeigte und sich negativ auf die Validität auswirken kann. Diese Begriffe wurden ersetzt oder umschrieben und Sätze verändert bzw. verkürzt. Diese Strategien zur Auswahl der Lösung wurden in diesem Beitrag anhand von einzelnen Beispielen aufgezeigt.

Durch die Identifikation dieser Strategien und Überprüfung der Verständlichkeit konnten für die Entwicklung von gebundenen Testaufgaben in der ökonomischen Bildung wertvolle Hinweise gewonnen werden. Die Entdeckung der Strategie, sich die ökonomische Bedeutung der Begriffe aus der Lebenswelt herzuleiten und bei ökonomischen Komposita diese vorab zu teilen, erlaubt einen Einblick in die Gedankenwelt der Schülerinnen und Schüler. Die Vorstellungen der Jugendlichen zu den einzelnen Fachbegriffen konnten bei dem gebundenen Aufgabenformat für die Gestaltung von falschen Antworten genutzt werden, um eher plausible, natürliche Antwortalternativen zu generieren und das Raten und einfache Ausschließen der falschen Antworten zu reduzieren. Sprachliche Verständnisprobleme konnten identifiziert und die Aufgaben in der Weise formuliert werden, dass sie in sinnvoller Weise von den Schülerinnen und Schülern verstanden wurden. Wenn nicht intendierte Prozesse reduziert werden und durch die Aufgaben eher der Wissende von dem Nichtwissenden unterschieden werden kann, dann kann auch eher davon ausgegangen werden, dass der Test das misst, was dieser messen soll.

4 Ausblick

Die Methode „Lauter Denken“ stellt eine Möglichkeit dar, Testaufgaben auf Itemebene hinsichtlich der Verständlichkeit und Güte zu untersuchen. Auch wenn diese qualitative Methode in dieser Weise durch die Durchführung, Transkription und Auswertung als eine eher

arbeitsintensive Methode bezeichnet werden kann, konnten durch die Analyse, hilfreiche Erkenntnisse über die Testaufgaben und den Lösungsprozess gewonnen werden, die bei der weiteren Konzeption von Items im Bereich der ökonomischen Bildung genutzt werden können. Allerdings stellen die im Rahmen der vorgestellten Studie erhobenen Kenntnisse nur einen ersten Einblick dar und sollten durch weitere Studien untermauert werden, um gesicherte und umfangreichere Ergebnisse erzielen zu können. Für die Entwicklung von gebundenen Aufgabenformaten zur Erfassung ökonomischer Kompetenzen könnte es sich zudem als nützlich erweisen, Items zuerst im offenen Aufgabenformat zu konzipieren und die Aufgaben unter Berücksichtigung der Methode „Lautes Denken“ lösen zu lassen. Die während der Methode „Lautes Denken“ verbalisierten Gedanken der Schülerinnen und Schüler könnten bei der Formulierung von Antwortalternativen bei gebundenen Aufgabentypen unterstützen. Durch dieses Vorgehen, das bereits auch in anderen Fachdidaktiken genutzt wird, könnten die Vorstellungen der Schülerinnen und Schüler zu den einzelnen ökonomischen Themeninhalten direkt als Distraktoren genutzt werden, um dadurch natürliche, plausible Antworten generieren zu können. Dieses Vorgehen wird für die Konzeption der Testaufgaben des Jahrgangs zehn in dieser Studie in Betracht gezogen.

Literaturverzeichnis

- Arndt, H./Kopp, B. (2016): Entwicklung eines Erhebungsinstruments zur Erfassung des Vorwissens von Grundschulkindern zu ökonomischen Präkonzepten – Ergebnisse aus der ersten Pilotierung. In: Arndt, H. (Hg.): Das Theorie-Praxis-Verhältnis in der ökonomischen Bildung. Schwalbach/Ts, 201-211.
- Beck, K./Krumm, V./in Zusammenarbeit mit Dubs, R. (1998): Wirtschaftskundlicher Bildungstest (WBT), Göttingen.
- Bernholt, S./Eggert, S./Kulgemeyer, C. (2012): Capturing the Diversity of Student's Competence in Science Classrooms: Differences and Commonalities of Three Complementary Approaches. In: Bernholt, S./Neumann, K./Nentwig, P. (Hg.): Making it tangible: Learning outcomes in science education, Münster, 173-199.
- Biggs, S. F./Rosman, A. J./Sergenian, G. K. (1993): Methodological issues in judgment and decision-making research – Concurrent verbal protocol validity and simultaneous traces of process. In: Journal of Behavioral Decision Making, 6, 187-206.
- Bise, V. (2008): Problemlösen im Dialog mit sich selbst. Dialogische Strukturen im inneren Sprechen beim Problemlösen – Eine explorative Studie nach der Methode des Lauten Denkens, Marburg.

- Buber, R. (2009): Denke-Laut-Protokolle. In: Buber, R./Holzmüller, H. H.: Qualitative Marktforschung Konzepte–Methoden–Analysen, Wiesbaden, 555-568.
- Bühler, K. (1907): Tatsachen und Probleme einer Psychologie der Denkvorgänge: I. Über Gedanken, Würzburg.
- Bühner, M. (2011): Einführung in die Test- und Fragebogenkonstruktion, 3., aktual. Aufl., München.
- Brandmaier, E./Frank, H./Korunka, C./Plessnig, A./Schopf, C./Tamegger, K. (2006): Ökonomische Bildung von Schüler/innen Allgemeinbildender Höherer Schulen – Arbeits-, Organisations- und Wirtschaftspsychologie 6, hg. v. Kirchner, E., Wien.
- Deffner, G. (1989): Interaktion zwischen Lautem Denken, Bearbeitungsstrategien und Aufgabenmerkmalen? Eine experimentelle Prüfung des Modells von Ericsson und Simon. In: Sprache und Kognition, 8, 1989, Heft 2, 98-111.
- Dresing, T./Pehl, T. (2011): Praxisbuch Transkription Regelsysteme, Software und praktische Anleitungen für qualitative ForscherInnen, 3. Aufl., Marburg.
- Englund Dimitrova, B. (2005): Expertise and Explicitation in the Translation Process. Amsterdam, Philadelphia.
- Ericsson, K. A./ Simon, H. A. (1980): Verbal reports as data. In: Psychological Review, 87, 215-251.
- Ericsson, K. A./ Simon, H. A. (1984): Protocol analysis – Verbal Reports as Data, Cambridge.
- Ericsson, K. A./ Simon, H. A. (1998): How to Study Thinking in Everyday Life: Contrasting Think Aloud Protocols With Descriptions and Explanations of Thinking. In: Mind, Culture, and Activity, 5, 3, 178-186.
- Flaherty, E. G. (1975): The Thinking Aloud Technique and Problem Solving Ability. In: The Journal of Educational Research, 1 February 1975, Vol. 68 (6), 223-225.
- Franzen, U./Merz, F. (1988): Einfluss des Verbalisierens auf die Leistung bei Intelligenzprüfungen – Neue Untersuchungen. In: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 8, 117-134.
- Frommann, U. (2005): Die Methode „Lautes Denken“. Online:http://www.e-teaching.org/didaktik/qualitaet/usability/Lautes%20Denken_e-teaching_org.pdf (20.07.2016).
- Funke, J./Spering, M. (2006): Methoden der Denk- und Problemlöseforschung. In: Funke, J. (Hg.): Denken und Problemlösen (Enzyklopadie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II. In: Kognition, Bd. 8, Göttingen, 647-744.
- Gausmann, E./Eggert, S./Hasselhorn, M./Watermann, R./Bögeholz, S. (2010): Wie verarbeiten Schüler/innen Sachinformationen in Problem- und Entscheidungssituationen Nachhaltiger Entwicklung. In: Klieme, E./Leutner, D./Kenk, M. (Hg.): Kompetenzmodellierung, Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. In: Zeitschrift für Pädagogik, 56, Beiheft, Weinheim und Basel, 204-215.
- Göpferich, S. (2007): Praktische Handreichung für Studien mit lautem Denken. Online: <http://www.susanne-goepferich.de/Handreichung.pdf> (20.07.2016).

- Greimel-Fuhrmann, B./Kronberger, R./Rumpold, H. (2016): Befunde und Desiderata zur Wirtschaftsbildung österreichischer Mittelschüler/innen. In: Wirtschaftspolitische Blätter 1/2016, 249-263.
- Hadenfeldt, J. C./Repinning, B./Neumann, K. (2014): Die kognitive Validität von Ordered Multiple Choice Aufgaben zur Erfassung des Verständnisses von Materie. In: Zeitschrift für Didaktik der Naturwissenschaften. doi:10.1007/s40573-014-0003-7, 57-68.
- Hartig, J./Frey, A./Jude, N. (2012): Validität. In: Moosbrugger, H./Kelava, A. (Hg.): Testtheorie und Fragebogenkonstruktion, 2., aktual. und überarb. Aufl., Berlin u. a., 143-172.
- Hentrich, S./Seeber, G. (2016): Wirtschaftskompetenzen von Schülerinnen und Schülern Ergebnisse einer Pilotstudie. Vortrag auf der Jahrestagung der Deutschen Gesellschaft für Ökonomische Bildung in Hildesheim am 24.-26.2.2016.
- Holsti, O. R. (1969): Content Analysis for the social and humanities, Reading, Mass.
- Jonskisz, E./Moosbrugger, H./Brandt, H. (2012): Planung und Entwicklung von Tests und Fragebogen. In: Moosbrugger, H./Kelava, A. (Hg.): Testtheorie und Fragebogenkonstruktion, 2., aktual. und überarb. Aufl., Berlin u. a., 27-74.
- Kirchner, V. (2016): Wirtschaftsunterricht aus der Sicht von Lehrpersonen – Eine qualitative Studie zu fachdidaktischen teachers' beliefs in der ökonomischen Bildung, Wiesbaden.
- Krell, M./Czeskleba, A./Krüger, D. (2012): Validierung von Forced Choice-Aufgaben durch lautes Denken. In: Erkenntnisweg Biologiedidaktik 11, 53-70.
- Knoblich, G./Öllinger, M. (2006): Die Methode des Lauten Denkens, The Method of Thinking Aloud. In: Funke, J./Frensch, P. A. (Hg.): Handbuch der Allgemeinen Psychologie – Kognition, Göttingen u. a., 691-696.
- Konrad, K. (2010): Lautes Denken. In: Mey, G./Mruck, K. (Hg.): Handbuch Qualitative Forschung in der Psychologie, Wiesbaden, 476-490.
- Kucan, L./Beck, I. L. (1997): Thinking Aloud and Reading Comprehension Research – Inquiry, Instruction, and Social Interaction. In: Review of Educational Research, Fall 1997, Vol.67, No 3, 271-299.
- Kulgemeyer, C. (2010): Physikalische Kommunikationskompetenz – Modellierung und Diagnostik, Berlin.
- Leuders, T./Naccarella, D./Philipp, K. (2011): Experimentelles Denken – Vorgehensweisen beim innermathematischen Experimentieren. In: Journal für Mathematik-Didaktik, 32 (2), 205-231.
- Loerwald, D./Schnell, C. (2014): Tests als Instrumente zur Individualdiagnostik in der ökonomischen Bildung. Konzeption, Validierung und Auswertung von Testaufgaben für die Sekundarstufe I in Niedersachsen. In: Retzmann, T. (Hg.): Ökonomische Allgemeinbildung in der Sekundarstufe I und Primarstufe, Schwalbach/Ts, 294-306.
- Niedersächsisches Kultusministerium (2013): Kerncurriculum für die Oberschule Schuljahrgänge 7-10 Wirtschaft, Hannover.
- Niedersächsisches Kultusministerium (2009a): Kerncurriculum für die Realschule Schuljahrgänge 8-10 Wirtschaft, Hannover.
- Niedersächsisches Kultusministerium (2009b): Kerncurriculum für die Hauptschule Schuljahrgänge 7-10 Wirtschaft, Hannover.

- Macha, K. (2015): Ökonomische Kompetenz messen – Theoretisches Modell und Ergebnisse der Economic Competencies Study (ECOS), Reihe: Ökonomische Bildung, Bd. 8, Berlin.
- Mackensen-Friedrichs, I. (2004): Förderung des Expertiseerwerbs durch das Lernen mit Beispielaufgaben im Biologieunterricht der Klasse 9. Dissertation zur Erlangung des Doktorgrades der Mathematisch-Naturwissenschaftlichen Fakultät der Christian-Albrechts-Universität zu Kiel, Kiel.
- Mayring, P. (2013): Qualitative Inhaltsanalyse. In: Flick, U./Kordoff, E. von/ Steinke, I. (Hg.): Qualitative Forschung – Ein Handbuch, 10. Aufl., Reinbek bei Hamburg, 468-474.
- Mayring, P. (2010): Qualitative Inhaltsanalyse Grundlagen und Techniken, 11., aktual. und überarb. Aufl., Weinheim und Basel.
- Prenzel, M./Walter, O./Frey, A. (2007): PISA misst Kompetenzen – Eine Replik auf Rindermann (2006): Was messen internationale Schulleistungsstudien? In: Psychologische Rundschau, 58 (2007) 2, 128-136.
- Rost, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. 2., vollst. überarb. und erw. Aufl., Bern u. a.
- Sandmann, A. (2014): Lautes Denken – die Analyse von Denk-, Lern- und Problemlöseprozessen. In: Krüger, D. u. a. (Hg.): Methoden in der naturwissenschaftsdidaktischen Forschung, Berlin/Heidelberg, 179-188.
- Schooler, J./Ohlsson, S./Brooks, K. (1993): Thoughts beyond words: When language overshadows insight. In: Journal of Experimental Psychology: General, 122, 166-183.
- Schumann, S./Eberle, F./Oepke, M. (2011): Über welche ökonomischen Kompetenzen verfügen Maturandinnen und Maturanden? Hintergrund, Fragestellungen, Design und Methode des Schweizer Forschungsprojekts OEKOMA im Überblick. In: Faßhauer, U./Aff, J./Fürstenau, B./Wuttke, E. (Hg.): Lehr-Lernforschung und Professionalisierung. Perspektiven der Berufsbildungsforschung, Farmington Hills, Schriftenreihe der Sektion Berufs- und Wirtschaftspädagogik der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE), 51-63.
- Schmidt-Atzert, L./Amelang, M. (2012): Psychologische Diagnostik, 5. Aufl., Berlin/Heidelberg.
- Schnell, R./ Hill, P. B./ Esser, E. (2011): Methoden der empirischen Sozialforschung, 9. Aufl., München.
- Silberer, G. (2005): Die videogestützte Rekonstruktion kognitiver Prozesse beim Ladenbesuch. In: Marketing ZFP, 27(4), 263-280.
- Völzke, K. (2012): Lautes Denken bei kompetenzorientierten Diagnoseaufgaben zur naturwissenschaftlichen Erkenntnisgewinnung. Zentrum für Lehrerbildung der Universität Kassel (Hg.): Reihe Studium und Forschung, Heft 20, Kassel.
- Weidle, R./Wagner A. C. (1982): Die Methode des Lauten Denkens. In: Huber, G. L./Mandl, H. (Hg.): Verbale Daten – Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung, Weinheim und Basel, 81-103.
- Wobker, I./Lehmann-Waffenschmidt, M./Kenning, P./Gigerenzer, G. (2012): What Do People Know About the Economy? In: Dresden Discussion Paper in Economics, No. 3/12.

- Würth, R./ Klein, H. J. (2001): Wirtschaftswissen Jugendlicher in Baden-Württemberg. Eine empirische Untersuchung, Künzelsau.
- Yom, M./Wilhelm, T. H./Gauert, S. (2009): Protokolle lauten Denkens und Site Covering. In: Buber, R./Holzmüller, H. H. (Hg.): Qualitative Marktforschung Konzepte – Methoden – Analysen, Wiesbaden, 637-652.