

## **Stimulated recall in der Sprachtestforschung. Ein praktisches Beispiel aus der Erprobung eines computerbasierten Leseverstehentests**

### **Abstract**

In diesem Beitrag wird über den Einsatz der Methode *stimulated recall* im Rahmen des Testforschungsprojekts *Task Lab* berichtet.<sup>1</sup> In diesem Projekt wurden Aufgabenmerkmale von computerbasierten Leseverstehensaufgaben untersucht. Der Fokus des vorliegenden Aufsatzes liegt auf der qualitativen Untersuchung dieser Aufgaben während der Erprobungsphase des ansonsten quantitativ ausgerichteten Projekts. Anhand von ausgewählten Datenauszügen wird dargestellt, welche Möglichkeiten der Einsatz von *stimulated recall* bei der Erprobung von Testaufgaben bietet. Auch werden die Grenzen der angewandten Methode zur Diskussion gestellt.

### **1. Einleitung**

Das Projekt *Task Lab* wurde 2014-2016 am Kompetenzzentrum für Mehrsprachigkeit (KFM) (Freiburg/Schweiz) durchgeführt<sup>2</sup>. Im Rahmen des Projekts wurden Leseverstehensaufgaben für Französisch als Fremdsprache erstellt und von Deutschschweizer Schüler/innen aus 35 sechsten Klassen gelöst. Die Studie war Teil der Vorbereitungsarbeiten zur „Überprüfung der Grundkompetenzen“ (ÜGK), bei welcher im Frühjahr 2017 erstmalig schweizweit das Erreichen der Lernziele in der ersten schulischen Fremdsprache am Ende der Primarstufe (6. Schuljahr) überprüft wurde. Das KFM war dabei mit der Entwicklung der computerbasierten Aufgaben der Fremdsprachentests in den Kompetenzbereichen Leseverstehen und Hörverstehen beauftragt<sup>3</sup>. Die Erfahrungen und Erkenntnisse aus *Task Lab* sollten wichtige Grundlagen für eine valide(re) Gestaltung der computerbasierten Aufgaben liefern. Es wurde aber auch erwartet, dass die Ergebnisse von *Task Lab* auch über die ÜGK hinaus genutzt werden könnten. Die Ausrichtung des Projekts war überwiegend quantitativ. Anlässlich der Erprobungsphase wurden aber mittels der an die Stimulated-Recall-Methode angelehnten Interviews mit 34 Schüler/innen auch qualitative Daten gewonnen. Über diese soll in diesem Beitrag berichtet werden.

---

<sup>1</sup> Die Autorin dankt Katharina Karges und Peter Lenz für ihre wertvollen Anmerkungen zum Manuskript des vorliegenden Textes.

<sup>2</sup> Mehr Informationen zum Projekt abrufbar unter: <http://www.institut-mehrsprachigkeit.ch/de/content/task-lab-untersuchungen-zum-besseren-verstaendnis-und-zur-erhoehung-der-validitaet-von> (17.09.2017).

<sup>3</sup> Für beide Kompetenzbereiche musste jeweils ein Gesamttest von rund 100 Minuten Bearbeitungsdauer erstellt werden (13 Einzeltests à 7-8 Minuten). Die beiden Tests wurden für den Einsatz in vier Lehrplanregionen in vier verschiedene Sprachvarianten übersetzt: 1) Schulsprache Französisch – Zielsprache Deutsch, 2) Schulsprache Deutsch – Zielsprache Französisch, 3) Schulsprache Deutsch – Zielsprache Englisch, 4) Schulsprache Italienisch – Zielsprache Französisch.

Es ist unumstritten, dass die Testerprobungsphase (*pre-operational testing*) bei der Konstruktion und Validierung von Sprachtests eine äusserst<sup>4</sup> wichtige Rolle spielt (vgl. Kenyon & MacGregor 2012). Diese Phase kann am besten als „the thoughtful, deliberate, and systematic process of collecting and analyzing evidence to support the validity of an assessment *prior* to that assessment’s operational use“ (Kenyon & MacGregor 2012: 295, Hervorhebung im Original) verstanden werden. Das Besondere an diesem Prozess ist, dass hier nicht etwa die Testteilnehmer/innen oder die Testanwender/innen im Fokus stehen, sondern der Test selbst. Trotz der Wichtigkeit dieser Phase für den Testentwicklungsprozess wird ihr bisher in der Testforschung erstaunlich wenig Aufmerksamkeit geschenkt. Kenyon und MacGregor (2012: 305) setzen sich daher zu Recht dafür ein, auch diesem Bereich der Sprachtestforschung die dringend nötige Aufmerksamkeit zu schenken:

We urge test developers to present and publish work on pre-operational testing, and the field as a whole to give it due regard as an important research enterprise. We hope that in five to ten years’ time there will be debates in the academic literature weighing the pros and cons of alternate approaches and discussing the most efficient and effective methods, so that all testers will understand the issues surrounding this overlooked topic and it will no longer be the underappreciated step-child of the test development process.

Im vorliegenden Beitrag soll diesem wichtigen Desiderat entsprochen werden. Am Beispiel des Projekts *Task Lab*, das im Folgenden näher dargestellt wird, wird aufgezeigt, wie *stimulated recall* als Datenerhebungstechnik bei der Erprobung von Sprachtests und Sprachtestaufgaben eingesetzt werden kann, welche wichtigen Erkenntnisse damit gewonnen werden können und was dabei zu beachten ist. Auch werden die Grenzen der gewählten Vorgehensweise diskutiert.

## 2. Projekt *Task Lab*

Das Ziel des Projekts *Task Lab* bestand in der Untersuchung des Einflusses ausgewählter Aufgabenfaktoren, was letztlich zu einer zuverlässigeren Interpretation der Testergebnisse und -skalen führen sollte. Verallgemeinerbare Erkenntnisse sollten zudem die Erstellung zukünftiger Testaufgaben unterstützen und deren Validität verbessern. In erster Linie wurde bei *Task Lab* untersucht, wie Aufgaben mit unterschiedlichen Formaten (d. h. Multiple Choice, Kurzantwort und Matching<sup>5</sup>) und mit unterschiedlichen sprachlichen Merkmalen (vor allem bezüglich der Sprache der Fragen und Antworten) funktionieren.

An der Hauptstudie nahmen im Frühjahr 2015 über 500 Schüler/innen aus rund 35 Klassen teil. Es handelte sich dabei um Sechstklässler/innen aus der Deutschschweiz,

---

<sup>4</sup> Der Beitrag folgt der Schweizer Rechtschreibung, in der das ß nicht verwendet wird.

<sup>5</sup> Zuordnungsaufgaben, bei denen aufgrund einer inhaltlichen Frage einer von 5-7 kurzen Texten ausgewählt werden musste. Zu den gleichen 5-7 Texte wurden je drei inhaltliche Fragen gestellt.

die seit der dritten Klasse Französisch als erste Fremdsprache<sup>6</sup> gelernt hatten (zwei bis drei Wochenlektionen<sup>7</sup>). Gemäss dem für sie geltenden Lehrplan sollten sie am Ende der sechsten Klasse Niveau A2.1 des Gemeinsamen Europäischen Referenzrahmens (Europarat 2001) bzw. des für diese Zielgruppe eingesetzten Europäischen Sprachenportfolios (Bersinger, Jordi, & Tchang 2005) erreicht haben. Jeder getestete Schüler bzw. jede getestete Schülerin arbeitete während insgesamt ca. 2,5 Lektionen<sup>8</sup> mit Instrumenten aus dem Projekt *Task Lab*, und zwar mit a) Aufgaben zum Leseverstehen (1 Lektion); b) einem Fragebogen (ca. 0,3 Lektionen); c) schriftlichen Aufgaben zu kognitiven und sprachlichen Kompetenzkomponenten (z. B. Wortschatz) (ca. 0,7 Lektionen); d) mündlichen Aufgaben zu kognitiven und sprachlichen Kompetenzkomponenten (0,5 Lektionen). Alle eingesetzten Erhebungsinstrumente wurden während des Feldtests von insgesamt 131 Schüler/innen aus neun Klassen erprobt. Bis auf den kurzen Papierfragebogen nach dem Leseverstehentest wurden alle Tests am Computer durchgeführt.

Im Rahmen des Projektes wurden insgesamt 18 Leseverstehensaufgaben (54 Items, d. h. je drei Items pro Aufgabe bzw. Lesetext) für Französisch als Fremdsprache entwickelt. Bei sechs der 18 Aufgaben handelte es sich um Matching-Aufgaben, die in zwei Sprachvarianten vorlagen (d. h. mit den gleichen Items einmal auf Deutsch und einmal auf Französisch). Bei den restlichen zwölf Aufgaben wurden für die gleichen Items sowohl die Sprache der Fragen und Antworten als auch das Antwortformat variiert, sodass alle Items in vier Versionen vorlagen: 1) Kurzantwort-Aufgaben auf Französisch, 2) Kurzantwort-Aufgaben auf Deutsch, 3) Multiple-Choice-Aufgaben auf Französisch und 4) Multiple-Choice-Aufgaben auf Deutsch. Die Testhefte, welche die Schüler/innen bearbeiteten, umfassten jeweils ungefähr gleich viele Aufgaben in jeder Sprachversion. Jeder Lesetext kam im Testheft nur einmal vor, d. h. nur in einer Sprachvariante. Insgesamt löste jedes Kind 13 Aufgaben: drei Matching- sowie zehn Multiple-Choice/Kurzantwort-Aufgaben.

### **3. Die Testerprobungsphase bei der Entwicklung von Sprachtests: Wann und wozu kann *stimulated recall* eingesetzt werden?**

Während einer Testerprobungsphase werden – je nach Fokus und Erkenntnisinteresse – sowohl qualitative als auch quantitative Erhebungsmethoden angewandt (vgl. Kenyon & MacGregor 2012). Bei der Erprobung von Sprachtestaufgaben wird mit qualitativen Methoden in der Regel eher in der ersten, erkundenden Etappe gearbeitet. In dieser Etappe, auch Pilotstudie (*pilot testing*) genannt, werden in der Regel mit

---

<sup>6</sup> Mehr Informationen zum Fremdsprachenunterricht in der Schweiz abrufbar unter: <http://www.edk.ch/dyn/11911.php> (08.05.2017).

<sup>7</sup> Drei Wochenstunden während der beiden ersten Lernjahre und zwei Wochenstunden im dritten und vierten Lernjahr.

<sup>8</sup> Jede Lektion dauerte 45 Minuten.

einer kleineren Probandenzahl qualitative Daten erhoben, um das Funktionieren des Tests besser kennenzulernen und seine eventuellen Schwachstellen<sup>9</sup> rechtzeitig identifizieren und beheben zu können. Quantitative Erhebungsmethoden werden hingegen generell eher beim Feldtest (*field testing*) angewandt, der in einer späteren Etappe der Testerprobung und mit einer grösseren, repräsentative(re)n Probandenzahl durchgeführt wird. Dabei wird vor allem untersucht, inwiefern die Items (psychometrisch) wie gewünscht funktionieren. Weil beim Einsatz von qualitativen und quantitativen Methoden zum Teil auch unterschiedliche Informationen erfasst werden, empfiehlt sich bei der Testerprobung laut Kenyon und MacGregor die Kombination der beiden methodischen Ansätze: „a mix of qualitative and quantitative research methods (sometimes known as „mixed methods“) can often provide a richer source of information than any single method alone“ (2012: 303).

Im Zentrum der ersten, explorativen Erprobung von Sprachtests stehen also typischerweise Fragen zu einzelnen Aufgaben und Items<sup>10</sup> sowie zum Test als Ganzem. Es wird dabei beispielsweise untersucht, ob die Richtlinien und Anweisungen zum Bearbeiten des Tests für alle Beteiligten klar sind, ob die Distraktoren (d. h. die nicht korrekten Multiple-Choice-Antwortmöglichkeiten) zwar plausibel, aber dennoch eindeutig falsch sind, oder ob bestimmte Items konstruktirrelevante Elemente beinhalten (wenn sie z. B. nur aufgrund des Weltwissens beantwortet werden können) (vgl. Kenyon & MacGregor 2012: 304).

Wichtige Einblicke, die zur Beantwortung dieser und ähnlicher Forschungsfragen beitragen, können mithilfe von introspektiven Datenerhebungsverfahren<sup>11</sup> gewonnen werden. Dabei handelt es sich um Verfahren, bei denen „gezielt Daten bezüglich einer bestimmten (mentalen oder interaktionalen) Tätigkeit erhoben werden“, und zwar durch lautes Aussprechen der Gedanken, Beobachtungen und Emotionen, welche im Zusammenhang mit dieser Tätigkeit stehen (Heine & Schramm 2016: 173)<sup>12</sup>.

Zu den introspektiven Datenerhebungsverfahren, die bei der Testerprobung angewandt werden können, gehört u.a. *stimulated recall*<sup>13</sup>. Dieses Verfahren wird von Gass und Mackey folgendermassen definiert:

(...) stimulated recall methodology is a technique in which participants are asked to recall thoughts they had had while performing a prior task or while they had participated in a prior event. It is assumed that some tangible (perhaps visual or aural) reminder of the event will stimulate recall of the mental processes in operation during

<sup>9</sup> Zum Beispiel unverständliche Anweisungen, Layout, Elemente der Testaufgaben, Probleme bei der Testadministration oder Bewertung.

<sup>10</sup> Nach Hasselgreen und Caudwell (2016: 53) verstehe ich unter Aufgaben bzw. Tasks „something we ask our learners to do, in order to show their knowledge or ability“ und unter Items „text-plus-task“. Im Projekt *Task Lab* gehörten zu je einer Aufgabe drei Items.

<sup>11</sup> Mehr zum Einsatz von introspektiven Datenerhebungsverfahren in der Testforschung u. a. bei Arras 2013; Cohen 2007; Green 1998; Lazaraton 2002; Lazaraton & Taylor 2007; Lumley & Brown 2005.

<sup>12</sup> Mehr zu den introspektiven Datenerhebungsverfahren sowie ihren Vor- und Nachteilen u. a. bei Banerjee 2004; Bowles 2010; Gass & Mackey 2007, 2017; Heine & Schramm 2016.

<sup>13</sup> In der deutschsprachigen Fachliteratur hat sich für eine Form von *stimulated recall* der Begriff „Lautes Erinnern“ etabliert (Knorr 2013, vgl. auch Heine & Schramm 2016). In diesem Artikel behalte ich jedoch die englischsprachige Terminologie bei.

the event itself and will, in essence, aid the participant in mentally reengaging with the original event. (2017: 14)

Die Methode des *stimulated recall* wird beispielsweise zur Erforschung von Lesestrategien (Barry & Lazarte 1998; Finkbeiner 2005; Finkbeiner, Knierim, Smasal, & Ludwig 2012; Gass & Mackey 2017) und von *test taking strategies* eingesetzt (Barkaoui, Brooks, Swain, & Lapkin 2013; Cohen 2007; Gass & Mackey 2017). Der Vorteil der Stimulated-Recall-Interviews gegenüber den retrospektiven Interviews liegt darin, dass erstere (insbesondere wenn zwischen der Aufgabe und dem Interview max. 24 Stunden liegen) weniger stark vom Erinnerungsvermögen der Proband/innen abhängig sind, da der Stimulus (z. B. ein Videoausschnitt oder ein Schreibprodukt) als Gedächtnisstütze und Ausgangspunkt für das Gespräch fungieren (Gass & Mackey 2017). Gegenüber Laut-Denken-Protokollen hat *stimulated recall* auch einen forschungspraktischen Vorteil: “For think-alouds the researcher needs to train participants, and even after training, not all participants are capable of carrying out a task and simultaneously talking about the task” (Gass & Mackey 2017: 15f.). Dieser Punkt ist vor allem in einem Forschungskontext, in dem der Zugang zum Feld schwierig und nicht selten auch zeitlich sehr begrenzt ist (wie es bei *Task Lab* der Fall war), nicht zu unterschätzen.

Aus den oben genannten Gründen wurde bei der Pilotierung der in *Task Lab* erstellten Leseverstehensaufgaben in Anlehnung an die Stimulated-Recall-Methode gearbeitet. Das konkrete Vorgehen wird im Folgenden näher beleuchtet.

#### **4. *Stimulated Recall* im Projekt *Task Lab***

Die im Projekt *Task Lab* durchgeführten Interviews zielten grundsätzlich darauf ab, zu ergründen, wie junge Französischlernende beim Lösen von computerbasierten Leseverstehensaufgaben vorgehen und welche Strategien sie dabei anwenden. Im Fokus stand die Frage nach dem Einfluss der Antwortformate und der Sprache der Fragen und Antworten auf das Testverhalten der Schüler/innen. Ein allgemeineres, eng mit der Validität der Aufgaben verbundenes Anliegen bestand zudem darin, herauszufinden, ob die Schüler/innen aus den „richtigen Gründen“ zu ihrer Lösung gelangen: Würde sich beispielsweise zeigen, dass die Schüler/innen eine richtige Antwort finden, ohne den Text gelesen zu haben, müsste das Testitem neu geschrieben werden. Darüber hinaus sollte auch ergründet werden, ob die Aufgaben und Items von Design und Technik her funktionierten bzw. was daran ggf. überarbeitet werden müsste. Die Aussagen der Schüler/innen wurden also nicht nur für die Validierung der Testinstrumente genutzt, sondern auch zur Überprüfung der *usability*, also der Benutzerfreundlichkeit.

Die Interviews wurden während der Erprobungsphase mit insgesamt 34 Schüler/innen aus zwei Klassen durchgeführt. Insgesamt vier Interviewer/innen<sup>14</sup> leiteten die Interviews, nachdem sie geschult worden waren. Die Proband/innen lösten

---

<sup>14</sup> Inklusive der Autorin des vorliegenden Textes.

während insgesamt 45 Minuten eine kleine Auswahl aus dem oben beschriebenen Aufgabenpool (üblicherweise 2-3 Aufgaben mit jeweils 3 Items) und gaben direkt nach jedem bearbeiteten Item Auskunft zu ihrem Vorgehen, ihren Überlegungen, sowie zu den eingesetzten Strategien und eventuellen Schwierigkeiten beim Lösen der Aufgaben.

Am Anfang einer Sitzung wurden jedem Kind die Ziele und der Ablauf der Untersuchung erklärt. Während das Kind las, zog sich die Interviewerin oder der Interviewer zurück, um es bei der Aufgabenbearbeitung nicht zu stören. Immer wenn das Kind signalisierte, dass es mit der Bearbeitung eines Items fertig war, wurde es dazu befragt (ca. 1-3 Minuten pro Item). Als Stimulus diente dabei die auf dem Bildschirm dargestellte Aufgabe und die vom Kind bei Matching und Multiple Choice ausgewählte bzw. beim Kurzantwortformat eingetippte Antwort. Die Gespräche fanden auf Deutsch und Schweizerdeutsch<sup>15</sup> statt.

Im Fokus der Interviews stand, wie bereits erwähnt, der Versuch einer Rekonstruktion der Lösungswege und Gedanken, welche die Schüler/innen zu ihren Antworten führten. Da es für die Zwecke des hier beschriebenen Projekts vor allem wichtig war, innerhalb kürzester Zeit möglichst viel über das Funktionieren der Testaufgaben zu erfahren, wurden den Schüler/innen nach Bedarf vertiefende Fragen gestellt. So wurden die Schüler/innen beispielsweise dazu befragt, ob sie die Frage auf Französisch verstanden haben, wie und wie oft sie den Text gelesen haben, wo im Text die Antwort gefunden wurde, ob die Aufgabenstellung klar war usw. Bei Multiple-Choice- und Matching-Items wurde auch gefragt, warum diese bestimmte Antwort ausgewählt wurde, und beim Format Kurzantwort, ob es evtl. Formulierungsschwierigkeiten gab. Während der letzten fünf Minuten eines Interviews wurden noch abschliessende und allgemeinere Fragen gestellt, z. B. ob es die Schüler/innen gestört hat, die Aufgaben am Computer zu lösen, ob das Navigationsprinzip für sie klar war, ob sie evtl. beim Schreiben auf der Tastatur Probleme mit französischen Sonderzeichen hatten und wie sie es fanden, dass bei gewissen Aufgaben der Text auf Französisch war und die Antworten auf Deutsch ausgewählt bzw. geschrieben werden mussten. Die Interviewer/innen stützten sich beim Gespräch auf einen schriftlichen Interviewleitfaden, der sich an den Forschungsfragen orientierte. Sie hielten zudem vor und während der Gespräche die Antworten der Schüler/innen sowie möglicherweise interessante Beobachtungen auf Ausdrücken der Leseverstehensaufgaben fest. Die Gespräche mit den Schüler/innen wurden aufgenommen. Später wurden sie transkribiert und mithilfe von MAXQDA kodiert. Die Auswertung fand nach den Prinzipien der Inhaltsanalyse (Mayring 2010) statt, wobei die Kategorien datengeleitet gebildet wurden.<sup>16</sup> Die Resultate dieser

---

<sup>15</sup> Die Schüler/innen konnten frei wählen, ob sie in ihrer Alltagssprache Schweizerdeutsch oder in der Schulsprache Standarddeutsch sprechen wollten.

<sup>16</sup> Für die Zwecke des hier beschriebenen Projekts war es dabei vor allem wichtig, möglichst viel über die Aufgaben selbst zu erfahren, daher wurde bei der Datenauswertung nicht unterschieden, bei welchen der erhobenen Daten es sich um erinnerte und bei welchen um im Nachhinein entstandene Gedanken handelte.

Datenanalyse bildeten die Basis für die Überarbeitung und Anpassung der eingesetzten Aufgaben zum Leseverstehen.

Generell konnte beobachtet werden, dass die meisten Schüler/innen sehr gerne Auskunft über ihre Gedanken und ihre Vorgehensweise gaben. Bei den Interviews entstand auch der Eindruck, dass die Schüler/innen über ihre Vorgehensweise beim Lösen der Testaufgaben generell sehr ehrlich berichtet haben: Sie gaben beispielweise problemlos zu, bei den Kurzantwort-Aufgaben vom Lesetext abgeschrieben oder beim Multiple Choice geraten zu haben. Die Tatsache, dass die Leistung im Test keinen Einfluss auf ihre Schulnote hatte (was den Kindern im Voraus so kommuniziert wurde), spielte dabei wohl eine wichtige Rolle.

## 5. Ausgewählte Resultate der Stimulated-Recall-Interviews

Im Folgenden soll am Beispiel einiger ausgewählter Resultate bezüglich der Sprache der Fragen und Antworten bei Fremdsprachentests (Kap. 5.1), der Erkennung missverständlicher Aufgabenstellungen und Textstellen (Kap. 5.2) sowie zur Suche nach Kognaten als Strategie (Kap. 5.3) das Potenzial von *stimulated recall* als Datenerhebungstechnik aufgezeigt und diskutiert werden.

### 5.1 Sprache der Fragen und Antworten<sup>17</sup>

Wie bereits erwähnt lagen bei *Task Lab* die gleichen Aufgaben mit den Fragen und Antworten in der Schulsprache Deutsch und in der Zielsprache Französisch vor. Die Stimulated-Recall-Interviews zeigten sehr deutlich auf, dass ungefähr die Hälfte der auf Französisch gestellten Fragen und beim Multiple-Choice-Format auch viele der Antwortoptionen von den Schüler/innen nicht oder falsch verstanden wurden. Einige Schüler/innen gaben an, den Text eigentlich gut verstanden zu haben, nur die französische Frage nicht. Das führte dazu, dass die Schüler/innen nicht wussten, ob sie die Frage richtig interpretiert hatten und ob ihr Ansatzpunkt für die Suche nach der richtigen Antwort im Text richtig war.

In den Interviews zeigte sich auch, dass Lernende, die eine Frage überhaupt nicht verstanden hatten, bei Multiple-Choice- und Matching-Items nicht selten einfach irgendeine Antwort ankreuzten und bei den Kurzantwort-Items irgendeine Antwort eingaben oder auch einfach aus dem Lesetext abschrieben.

S<sup>18</sup>: Ich habe zuerst die Frage gelesen, die ich aber nicht ganz verstanden habe. Dann habe ich versucht, sie zu verstehen. Und dann wieder den Text lesen. (...) Und weil ich die Frage nicht ganz verstanden habe, musste ich etwas raten.

Bei manchen Fragen bzw. Antwortoptionen konnten die Kinder gar nur einzelne (mit dem Deutschen verwandte) Wörter entschlüsseln.

*Multiple-Choice-Frage zum Text:*

---

<sup>17</sup> Mehr Informationen dazu in Barras, Karges, & Lenz 2016.

<sup>18</sup> S=Schüler/in

*Quel est le problème avec C dans cette scène?*

a) *A et B doivent faire leurs devoirs.*

b) *A et B ne doivent pas parler.*

c) *A et B ont perdu un livre.*<sup>19</sup>

S: Ich verstehe keine einzige Antwort.

I<sup>20</sup>: Wie lautet die Frage?

S: Wer löst die Probleme in dieser Szene. (...) Ich glaube, es ist das dritte. (...)

I: Ok, aber verstehen tust du's eigentlich nicht.

S: Ich glaube, „livre“ heisst... Also A und B... Nein.

Es konnte aber auch beobachtet werden, dass in einigen Fällen der Sinn einer Frage nur wegen einem einzelnen, nicht oder falsch verstandenen Wort falsch gedeutet wurde.

*Frage zum Text:*

*Qui préfère les langues aux maths ?*<sup>21</sup>

I: Was heisst es?

S: Welche Sprachen sie in Mathe sprechen.

Auch im Zusammenhang mit dem Kurzantwort-Format deckten die Interviews deutliche Probleme auf: Wenn die Frage auf Französisch gestellt wurde, musste sie auch in dieser Sprache beantwortet werden. Das Schreiben auf Französisch wurde von vielen Kindern aber als schwierig empfunden:

I: Was sagst du dazu, dass du auf Französisch schreiben musstest?

S: Es ist ja ziemlich schwierig, weil man zusätzlich überlegen muss, wie man etwas schreibt.

Ausserdem genügten ihre Wortschatzkenntnisse oft nicht, um auch nur eine sehr kurze Antwort formulieren zu können.

Die Interviews haben auch gezeigt, dass die Schüler/innen manchmal etwas ganz Anderes schreiben wollten, als sie tatsächlich auf Französisch geschrieben haben.

I: Kannst du mir deine Antwort bitte vorlesen?

S: Le job est régulièrement.<sup>22</sup>

I: Was heisst das auf Deutsch?

S: Der Job ist unregelmässig.

S: Ich habe geschrieben: „Pierre Dumont en danger.“<sup>23</sup> (...)

<sup>19</sup> „Was ist das Problem mit C in dieser Szene? a) A und B sollen ihre Hausaufgaben machen. b) A und B sollen nicht sprechen. c) A und B haben ein Buch verloren.“

<sup>20</sup> I=Interviewer/in

<sup>21</sup> „Wer mag lieber Sprachen als Mathematik?“

<sup>22</sup> „Der Job ist regelmässig.“

<sup>23</sup> Pierre Dumont in Gefahr“.



I: Was würdest du auf Deutsch schreiben, wenn du diese Antwort auf Deutsch schreiben könntest?

S: Pierre Dumont ist sehr gefährlich.

Erkenntnisse wie diese könnten wiederum erklären, warum bei der Auswertung der Hauptstudie ca. ein Drittel der Kurzantworten auf Französisch als „abwegig“ klassifiziert werden mussten, weil zwischen der Frage und der Antwort kein inhaltlicher Zusammenhang erkennbar war.

Es gab aber auch Kinder, welche die Fragen in der französischen Version bevorzugten. Manche von ihnen konnten bereits gut Französisch, weil sie diese Sprache entweder zu Hause oder mit einem Teil der Familie sprachen. Andere gaben in den Gesprächen an, dass sie die französischen Fragen einfacher fanden, weil sie dann die Stelle im Text, auf die sich die Frage bezog, einfacher identifizieren konnten (allerdings waren die Fragen so konzipiert, dass diese Strategie oft nicht aufgehen konnte). In einem Interview wurde erwähnt, dass es schwieriger sei, die Fragen auf Deutsch zu beantworten, wenn man den Lesetext nicht verstanden hatte. Argumente wie diese sprechen gerade *gegen* die Formulierung von Fragen in der Zielsprache, wenn das Ziel die Überprüfung des Textverständnisses in der Zielsprache ist.

Die Resultate der Interviews mit den Schüler/innen zeigen klar, dass die Leseverstehensitems mit Fragen und Antworten in der Schulsprache zu valideren Ergebnissen geführt haben als solche in der Zielsprache. Auf dieser Grundlage konnte folgende Empfehlung<sup>24</sup> für Tests in Kontexten mit einer gemeinsamen Schulsprache formuliert werden:

Beim Testen von fremdsprachlichen Kompetenzen im Leseverstehen sollten Fragen und Antworten in der Schulsprache vorgesehen werden. Mithilfe von Fragen und Antworten in der Schulsprache kann zuverlässiger das gemessen werden, was auch tatsächlich gemessen werden soll: die Fähigkeit, zielsprachliche Texte zu verstehen. Wenn dafür die Zielsprache verwendet wird, besteht die Gefahr, dass die Schülerinnen und Schüler an nicht oder falsch verstandenen Fragewörtern, unvollständigem Verstehen von Antwortoptionen oder an ungenügenden Schreibfertigkeiten scheitern, anstatt dass sie ihre Kompetenzen in dem Bereich unter Beweis stellen können, um den es im Test gehen soll: dem Textverstehen (Barras, Karges & Lenz 2016: 18).

## **5.2 Erkennen missverständlicher Aufgabenstellungen und Textstellen**

Die Erkenntnis, dass die Formulierung der Testfragen eine äusserst wichtige Rolle spielt, ist in der Testforschung bestens bekannt. Die Stimulated-Recall-Interviews zeigten auf, welche Formulierungen bei den eingesetzten Aufgaben zu Missverständnissen führten und deswegen angepasst werden mussten.

So zeigten die Gespräche mit den Schüler/innen beispielsweise, dass die ursprünglichen Du-Formulierungen in Matching-Aufgaben abgeändert werden

---

<sup>24</sup> Diese Empfehlung wurde auch in den Fremdsprachen-Aufgaben der ÜGK (schweizerisches Bildungsmonitoring) umgesetzt.

mussten, weil die Schüler/innen solche Aufgabenstellungen irrtümlicherweise auf sich selbst bezogen und die Fragen nicht aufgrund der Lesetexte, sondern ihrer eigenen Vorlieben beantworteten:

*Matching-Aufgabe: Links siehst du fünf kurze Aufgabenstellungen aus der Schule*<sup>25</sup>.

*Dazu stellen wir dir drei Fragen.*

*Frage 1: Wo brauchst du ein Wörterbuch?*

I: Wie bist du vorgegangen?

S: Ich habe alle durchgelesen und habe geschaut, wo ich am wenigsten verstehe. Bei diesen beiden war ich nicht ganz sicher, aber bei dem habe ich etwas mehr verstanden.

I: Wonach hast du gesucht? Welchen Text hast du ausgewählt? (...)

S: Den, den ich am wenigsten verstanden habe.

I: Hast du die Frage auch gelesen?

S: Ja, wo brauchst du ein Wörterbuch.

I: Aha, du hast es so verstanden? Du hast verstanden, für welchen Text DU ein Wörterbuch brauchst?

S: Ja.

Das Problem konnte gelöst werden, indem Formulierungen wie „Wo brauchst du ein Wörterbuch?“ durch „Wo braucht man ein Wörterbuch?“ ersetzt wurden.

Dank der Interviews konnte auch eine Reihe von sprachlichen Elementen identifiziert werden, über welche die Schüler/innen beim Lösen der Aufgaben stolperten. Beispielsweise wurde in der französischsprachigen Version einer Frage zum Text, der von einem Herrn namens Pierre Dumont handelte, die Abkürzung „M. Dumont“ für „Monsieur Dumont“<sup>26</sup> benutzt. Es zeigte sich, dass diese Abkürzung für manche Schüler/innen verwirrend war und deswegen im Lesetext ausgeschrieben werden musste:

S: Hä, was ist das da... M. Dumont? Das weiss ich jetzt nicht.

I: Ja, das ist in Ordnung. Was hast du jetzt nicht verstanden? Womit hast du Mühe?

S: Da hier steht M. und da hier steht Pierre.

In einem anderen Text, der in Form eines Chats präsentiert wurde, wurde in der Erprobungsphase die in Chatrooms übliche Abkürzung „@Timo“ benutzt. Es hat sich herausgestellt, dass nicht alle Kinder mit der Bedeutung des Symbols @ vertraut waren und deswegen nicht erkennen konnten, wer der Adressat und wer der Verfasser der Nachricht ist. Damit diese konstruktirrelevante (d.h. nichts über die Leseverstehenskompetenz aussagende) Kleinigkeit die Schüler/innen nicht verwirrt, wurde das Symbol anschliessend aus den Lesetexten entfernt.

Erkenntnisse wie diese zeigen, dass *stimulated recall*, obwohl zeitintensiv, bei der Erstellung von Sprachtests mit Gewinn eingesetzt werden kann. Es ist anzunehmen,

<sup>25</sup> Der gesuchte Text lautete: „Fais une affiche avec des mots que tu trouves intéressants. Ecris les mots que tu connais sur l’affiche et ensuite, cherche d’autres mots dans le dictionnaire.“ (*Mach ein Plakat mit Wörtern, die du interessant findest. Schreib die Wörter, die du kennst, auf ein Plakat und such dann im Wörterbuch nach weiteren Wörtern*).

<sup>26</sup> Herr Dumont.

dass dies auch bei der Erprobung von neuen Unterrichtsmaterialien, Aufgabenstellungen, Aufgabentypen, und Unterrichtsformen der Fall sein könnte.

### 5.3 Suche nach Kognaten als Strategie

Generell konnte bei den Interviews beobachtet werden, dass die meisten Kinder sehr mutig an die französischen Texte herangingen. Sie liessen sich nicht vom ersten unbekanntem Wort abschrecken und setzten z. T. sehr gezielt bestimmte Strategien ein, um die Textbedeutung zu entziffern. Eine beliebte Strategie, welche die Schüler/innen in den Stimulated-Recall-Interviews häufig erwähnten, ist die Suche nach Kognaten, also etymologisch verwandten Wörtern, in den Texten, Fragen und ggf. Antwortoptionen.

I: Wie hast du den Text gelesen?

S: So, wie er da steht. Einfach durchgelesen. Ein paar Parallelwörter (d. h. Kognaten, Anm. M.B.) und das, was wir bereits gelernt haben, habe ich verstanden und so konnte ich mir einen Reim daraus machen.

Der häufige Einsatz der Kognaten-Strategie ist für diese Schüler/innengruppe nicht ganz überraschend: Sie spielt im geltenden Lehrplan eine wichtige Rolle und wird dort an mehreren Stellen explizit genannt. So sollen die Schüler/innen beispielsweise im Kompetenzbereich *Lernstrategische Kompetenzen* in den Schuljahren 5/6 im Fach Französisch Folgendes lernen: „Lesen: Texte erschliessen durch Bilder-Anschauen und Titel-Beachten, Bekanntes und Parallelwörter in anderen Sprachen suchen, raten“ (Passepartout: Fremdsprachen an der Volksschule 2015: 25). Im Französisch-Lehrwerk *Mille feuilles*, das auf diesen Lehrplan abgestimmt ist, spielt die Strategie ebenfalls eine wichtige Rolle.

Die Gespräche und die Analyse der Testresultate zeigen, dass diese Strategie vor allem dann aufging, wenn die Fragen zum Text das globale Textverständnis überprüften. Dies war beispielsweise bei einem Lesetext der Fall, in dem ein Mädchen über ihren Tag im Zirkus berichtet und die Schüler/innen gefragt wurden, worum es im Text geht.

S: Also wir haben in der Schule oft für Parallelwörter geübt. Und die sind eben sehr ähnlich. Zum Beispiel Giraffe oder auch über einen Zirkus. Ich sehe da Akrobatik, das habe ich verstanden, und so konnte ich manche Wörter, die ich nicht verstanden habe, dann einreihen. So wusste ich einfach ungefähr das Thema, Details nicht ganz. So habe ich das meiste verstanden.

Auffallend war, dass die Schüler/innen Kognaten in vielen Fällen nicht nur vom Deutschen, sondern auch vom Englischen herleiteten, z. B. *histoire* von *history*, *vétérinaire* von *veterinarian*, *ennuyeux* von *annoying*. Das ist eine überraschende Erkenntnis, vor allem angesichts der Tatsache, dass diese Schüler/innen Englisch erst seit einem guten Jahr lernten, Französisch hingegen bereits zwei Jahre länger.

Diese Strategie stellte sich in manchen Fällen auch bei Detailfragen als hilfreich heraus. Oft führte sie jedoch zum Misserfolg. Dies hatte unterschiedliche Gründe. In manchen Fällen verstanden die Schüler/innen nur die Kognaten, den eigentlichen Sinn

des Textes konnten sie aber allein auf dieser Grundlage nicht erfassen. Einige Kinder konstruierten dafür rund um diese Wörter, die sie mehr oder weniger gut verstanden, zum Teil richtiggehend Fantasiegeschichten, die mit dem Lesetext nur noch marginal etwas zu tun hatten.

Es wurde auch festgestellt, dass die Schüler/innen in manchen Fällen französische Wörter falsch als Kognaten identifizierten und übersetzten. So wurde z. B. *travailler* (arbeiten) als „reisen“ (vom Englischen *travel*) verstanden, *plat préféré* (Lieblingsspeise) als „Lieblingspflanze“ (wohl vom Englischen *plant*), *cirque* (Zirkus) als „Zirkel“ bzw. „Kreis“, oder *beaucoup* (viel) als „schön“ (wegen Französisch *beau* = schön)<sup>27</sup>. Wurden die (vermeintlichen oder tatsächlichen) Kognaten falsch verstanden, führte es oft dazu, dass auf die entsprechenden Fragen falsche bzw. abwegige Antworten gegeben wurden. Dies war in rund der Hälfte aller Situationen der Fall, in denen in den Interviews der Einsatz von Kognatensuche als Strategie dokumentiert werden konnte.

Die hier dargestellten Beispiele bestätigen die aus der Strategieforschung bereits bekannte Tatsache, dass der Einsatz von Lese- bzw. Testbewältigungsstrategien nicht automatisch zum Erfolg führen muss:

Some learners might employ several strategies in order to solve a specific comprehension problem but without success either (1) because they lack the knowledge or skills necessary to actually apply the chosen strategies or (2) because the chosen strategies are not applicable to the resolution of the problem at hand (Finkbeiner et al. 2012: 60).

Die im Rahmen von *Task Lab* durchgeführten Interviews zeigen auf, wie wichtig es ist, zwar den Einsatz von Strategien im Unterricht zu thematisieren und zu üben, die Schüler/innen (und Lehrpersonen) aber auch darauf zu sensibilisieren, dass Strategien situationsadäquat eingesetzt werden müssen und auch die besten Strategien ohne genügende Sprachkenntnisse nicht immer zielführend sind. Der Erwerb von Wortschatz und Sprachstrukturen beim Erlernen einer Fremdsprache sollte daher auf keinen Fall zu kurz kommen.

## 6. Diskussion

Die Stimulated-Recall-Interviews im Projekt *Task Lab* erwiesen sich als wahre Datenfundgrube. Sie ermöglichten es, wichtige Einsichten in das Funktionieren der computerbasierten Leseverstehensaufgaben zu gewinnen und trugen erheblich zur Verbesserung der Aufgaben bei (vgl. Kap. 5.1 und 5.2). Auch können bestimmte Erkenntnisse für Fremdsprachenlehrpersonen und Entwickler/innen von Lehrmaterialien von Bedeutung sein (vgl. Kap. 5.3). Nichtsdestotrotz soll an dieser

---

<sup>27</sup> Ähnlich wurden auch „normale“ französische Wörter missverstanden. Kinder übersetzten beispielsweise *bulle* (Sprechblase) als „Hühnchen“ (wegen fr. *poulet* bzw. schweizerdeutsch „Poulet“), *cadeau* (Geschenk) als „Einkaufswagen“ (wegen fr. *caddie*) oder *boîte* (Dose) als „Holz“ (fr. *bois*). Ohne die mündlichen Erklärungen der Kinder würde man in manchen Fällen auch mit viel Vorstellungskraft nicht herausfinden, wie die Kinder auf ihre Antworten kamen.

Stelle auch auf gewisse Grenzen des angewandten Datenerhebungsverfahrens hingewiesen werden.

Die Auswertung der Gespräche hat gezeigt, dass die genaue Rekonstruktion der Lösungswege der Schüler/innen beim Bearbeiten der Leseverstehensaufgaben mithilfe von *stimulated recall* nicht immer ganz gelungen ist. Ein wichtiger Grund dafür dürfte sein, dass sich die retrospektiven Äusserungen zwar auf Überlegungen und Gedanken während der Handlung beziehen können, sie diese jedoch nicht direkt bzw. nicht vollumfänglich wiedergeben (vgl. Knorr 2013: 38). Bei Forschungsfragen, bei welchen die Lösungswege der Proband/innen im Zentrum stehen, wäre daher der Einsatz anderer Datenerhebungsmethoden (z. B. des Lauten Denkens) u. U. zielführender.

Die Schüler/innen berichteten in den Gesprächen nur von dem, was ihnen auch selber bewusst war, und worüber sie selber auch berichten wollten. So wurde zum Beispiel in manchen Fällen schnell deutlich, dass die Schüler/innen bei ihrer Antwort geraten hatten. Zwar ist dies an sich nicht uninteressant, die Gründe für das Nichtverstehen des Textes oder der Frage konnten aber nicht immer eruiert werden:

I: Warum hast du die erste Antwort gewählt?

S: Ich bin mir nicht immer ganz sicher. Ich rate auch ein bisschen. Ich weiss nicht, warum ich die gewählt habe.

Die Interviews konnten auch nur ansatzweise Hinweise dafür liefern, welchen Einfluss die Aufgabenformate (v.a. Multiple Choice und Kurzantwort) auf die Leseprozesse der Schüler/innen hatten. Es konnte aber beobachtet werden, dass das Vorgehen der Schüler/innen beim Lösen der Testaufgaben nicht bei allen Kindern auf die gleiche Art und Weise durch unterschiedliche Antwortformate beeinflusst war. Dies könnte ein Hinweis darauf sein, dass bei weiteren Untersuchungen das Spektrum der individuellen Lösungswege bei den einzelnen Items gezielter erfasst werden sollte.

Des Weiteren konnte bei der Auswertung der Transkripte beobachtet werden, dass die einzelnen Interviewer/innen die Gespräche zum Teil sehr unterschiedlich führten, und zwar trotz eines vorherigen Trainings und des Einsatzes eines einheitlichen Interviewleitfadens. So haben sie z. B. unterschiedlich hartnäckig nachgefragt oder zum Teil andere Fragen gestellt. Aus diesem Grund sind die Daten nicht immer gleich ausführlich und auch mussten gewisse Daten aus der Auswertung ausgeschlossen werden.

Auch wenn im Projekt *Task Lab* mit *stimulated recall* nicht allen Forschungsfragen in der gewünschten Ausführlichkeit nachgegangen werden konnte, waren die Interviews dennoch eine wichtige und reiche Informationsquelle und gewährten wertvolle Einblicke in die Lese- und Lösungsprozesse der Schüler/innen.

## 7. Fazit

Die in diesem Beitrag dargestellten Erkenntnisse zeigen, dass *stimulated recall* als Datenerhebungstechnik bei der Sprachtesterprobung, aber auch allgemein in der Unterrichtsforschung (z. B. beim Erproben neuer Unterrichtsmaterialien), durchaus zu empfehlen ist.

Stimulated-Recall-Interviews können wichtige Informationen liefern, die quantitative Forschungsmethoden allein nicht erzeugen könnten. Im Projekt *Task Lab* wurden mit dieser Technik wichtige Erkenntnisse über das Funktionieren der Leseverstehensaufgaben gewonnen, was zur Verbesserung ihrer Qualität und damit der Testvalidität beigetragen hat. Es kann vermutet werden, dass die hier dargestellten Ergebnisse auch für Fremdsprachendidaktiker/innen und Lehrmaterialienentwickler/innen von Interesse sein können.

## Literaturverzeichnis

- Aguado, Karen; Heine, Lena & Schramm, Karen (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung*. Frankfurt am Main: Peter Lang.
- Arras, Ulrike (2013), Introspektive Verfahren in der Sprachtestforschung. In: Aguado, Karen; Heine, Lena & Schramm, Karen (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung*. Frankfurt am Main: Peter Lang, 74–91.
- Banerjee, Jayanti (2004), Reference supplement to the preliminary pilot version of the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment. Section D: Qualitative analysis methods. Council of Europe, Language Policy Division (abrufbar unter: <http://www.coe.int/t/dg4/linguistic/CEF-ref-supp-SectionD.pdf>, 08.05.2017).
- Barkaoui, Khaled; Brooks, Lindsay; Swain, Merrill & Lapkin, Sharon (2013), Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34(3), 304–324.
- Barras, Malgorzata; Karges, Katharina & Lenz, Peter (2016), Leseverstehen überprüfen. Welche Sprache für die Fragen und Antworten in den Testitems? *Babylonia*, 2, 13–18.
- Barry, Sue & Lazarte, Alejandro A. (1998), Evidence for mental models: How do prior knowledge, syntactic complexity, and reading topic affect inference generation in a recall task for nonnative readers of Spanish? *The Modern Language Journal*, 82(2), 176–193.
- Bersinger, Simone; Jordi, Urs, & Tchang, Martine (2005). *Europäisches Sprachenportfolio. Version für Kinder und Jugendliche von 11 bis 15 Jahren (ESP II)*. Bern: Schulverlag.
- Bowles, Melissa A. (2010), *The think-aloud controversy in second language research*. New York: Routledge.
- Caspari, Daniela; Klippel, Friederike; Legutke, Michael K. & Schramm, Karen (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch*. Tübingen: Narr.
- Cohen, Andrew D. (2007), The coming of age for research on test-taking strategies. In: Fox, Janna D.; Welsche, Mari; Bayliss, Doreen; Cheng, Liying; Turner, Carolyn E. & Doe, Christine (Hrsg.), *Language testing reconsidered*. Ottawa: University of Ottawa Press, 89–111.
- Europarat (Hrsg.) (2001), *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Berlin, München: Langenscheidt.
- Finkbeiner, Claudia (2005), *Interessen und Strategien beim fremdsprachlichen Lesen: Wie Schülerinnen und Schüler englische Texte lesen und verstehen*. Tübingen: Narr.
- Finkbeiner, Claudia; Knierim, Markus; Smasal, Marc & Ludwig, Peter H. (2012), Self-regulated cooperative EFL reading tasks: Students' strategy use and teachers' support. *Language Awareness*, 21(1–2), 57–83.
- Fox, Janna D.; Welsche, Mari; Bayliss, Doreen; Cheng, Liying; Turner, Carolyn E. & Doe, Christine (Hrsg.), *Language testing reconsidered*. Ottawa: University of Ottawa Press.
- Fulcher, Glenn & Davidson, Fred (Hrsg.), *The Routledge handbook of language testing*. Milton Park, Abingdon, Oxon, New York: Routledge.
- Gass, Susan M. & Mackey, Alison (2007), *Data elicitation for second and foreign language research*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Gass, Susan M. & Mackey, Alison (2017), *Stimulated recall methodology in applied linguistics and L2 research* (Second edition). New York: Routledge.
- Green, Alison (1998), *Verbal protocol analysis in language testing research: A handbook*. Cambridge: Cambridge University Press.
- Hasselgreen, Angela & Caudwell, Gwendydd (2016), *Assessing the language of young*

- learners*. Sheffield, Bristol: Equinox Publishing Ltd.
- Heine, Lena & Schramm, Karen (2016), Introspektion. In: Caspari, Daniela; Klippel, Friederike; Legutke, Michael K. & Schramm, Karen (Hrsg.), *Forschungsmethoden in der Fremdsprachendidaktik: Ein Handbuch*. Tübingen: Narr, 173–181.
- Hinkel Eli (Hrsg.), *Handbook of research in second language teaching and learning*. New York: Lawrence Erlbaum Associates Inc.
- Kenyon, Dorry M. & MacGregor, David (2012), Pre-operational testing. In: Fulcher, Glenn & Davidson, Fred (Hrsg.), *The Routledge handbook of language testing*. Milton Park, Abingdon, Oxon, New York: Routledge, 295–306.
- Knorr, Petra (2013), Zur Differenzierung retrospektiver verbaler Daten: Protokolle Lauten Erinnerns erheben, verstehen und analysieren. In: Aguado, Karen; Heine, Lena & Schramm, Karen (Hrsg.), *Introspektive Verfahren und Qualitative Inhaltsanalyse in der Fremdsprachenforschung*. Frankfurt am Main: Peter Lang, 31–53.
- Kompetenzzentrum für Mehrsprachigkeit, *Projekt Task Lab* <http://www.zentrum-mehrsprachigkeit.ch/de/content/task-lab-untersuchungen-zum-besseren-verstaendnis-und-zur-erhoehung-der-validitaet-von> (abrufbar unter: 17.09.2017).
- Lazaraton, Anne (2002), *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.
- Lazaraton, Anne & Taylor, Lynda (2007), Qualitative research methods in language test development and validation. In: Fox, Janna D.; Welsche, Mari; Bayliss, Doreen; Cheng, Liying; Turner, Carolyn E. & Doe, Christine (Hrsg.), *Language testing reconsidered*. Ottawa: University of Ottawa Press, 113–129.
- Lumley, Tom & Brown, Annie (2005), Research methods in language testing. In: Hinkel Eli (Hrsg.), *Handbook of research in second language teaching and learning*. New York: Lawrence Erlbaum Associates Inc., 833–855.
- Mayring, P. (2010), *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim und Basel: Beltz.
- Passepartout: Fremdsprachen an der Volksschule (2015), *Lehrplan Französisch und Englisch* (abrufbar unter: <http://www.passepartout-sprachen.ch/services/downloads/>, 08.05.2017).
- Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK), *Sprachenunterricht*. Abrufbar unter: <http://www.edk.ch/dyn/11911.php> (08.05.2017)
- Taylor, Lynda (2005), Using qualitative research methods in test development and validation. *Research Notes*, (21), 2–4 (abrufbar unter <http://www.cambridgeenglish.org/images/23140-research-notes-21.pdf>, 08.05.2017).