

Lehrerhandbuch zum Lehrbuch →	
Hilfen zur Erstellung von Lernfortschrittstests Sibylle Bolton	

Bei Seminaren für den SDU bitten die Dozenten immer wieder um Hilfen für die Prüfungs- und Testerstellung. Deshalb wurden im Rahmen dieser Lehrerhandreichungen zu allen Sprachaktivitäten Prüfungsaufgaben mit Lösungsvorschlägen und Bewertungsanleitungen entwickelt.

Als Einstieg in dieses sehr komplexe Thema finden Sie hier den Artikel von Frau Dr. Sibylle Bolton (früher Prüfungsexpertin Goethe-Institut München, heute AUDIT-Expertin für Internationale Prüfungszentren).

Dieser Beitrag wurde für die Anhänge der Rahmencurricula für den SDU geschrieben und ist auch als Anhang 21 des Rahmencurriculum für den SDU in Belarus (Minsk 2010) aufgenommen worden (s. weitere Informationen zu den Rahmencurricula unter [2.2.2 SDU-CURRICULUM Bosnien und Herzegowina](#)).

Dieser Artikel beschreibt nicht nur die Gütekriterien für Tests, sondern zeigt auch, wo bei bestimmten Aufgabentypen die Fehlerquellen liegen.

Hilfen zur Erstellung von Lernfortschrittstests

Sibylle Bolton

1. Gütekriterien für Tests

Alle Arten von Tests, ob es sich nun um formelle Prüfungen oder Lernfortschrittstests handelt, sollten die Leistungen der Lernenden möglichst genau erfassen (inhaltliche Validität), möglichst zuverlässig messen (Reliabilität) und möglichst objektiv bewerten.

1.1 Inhaltliche Validität

Dieses Kriterium bedeutet, dass ein Test auch wirklich das prüft, was er prüfen soll. Die inhaltliche Validität wird durch einen sorgfältigen Abgleich der Prüfungsziele mit den Testaufgaben hergestellt. Wenn eine Testaufgabe zum Beispiel überprüfen soll, ob ein schriftlicher Text in seinen

Hauptaussagen und Einzelheiten verstanden wird, dann müssen die Aufgaben zum Text sowohl das Globalverstehen wie auch das Detailverstehen testen. Wenn eine Testaufgabe hingegen überprüfen soll, ob die Lernenden einem Text gezielt bestimmte Informationen entnehmen können, müssen die Aufgaben das selektive Verstehen testen. Die Voraussetzung für inhaltlich valide Testaufgaben zu den rezeptiven Fertigkeiten ist somit eine genaue Definition der Prüfungsziele, die deutlich macht, *welche Textsorten* verstanden werden sollen und *wie* diese Texte verstanden werden sollen.

1.2 Reliabilität

Die Reliabilität betrifft die Messgenauigkeit der Testaufgaben. Ein Test sollte möglichst so präzise messen wie ein Metermaß, das auch bei mehrmaliger Messung des gleichen Gegenstandes immer die gleichen Maße ergibt. In Wirklichkeit verhält es sich aber bei Tests nicht ganz so, da bei der „Messung“ sprachlicher Leistungen auch noch andere Faktoren als die sprachliche Leistung an sich eine Rolle spielen. Eine häufige Quelle von Fehlern bei der „Messung“ liegt zum Beispiel in den Arbeitsanweisungen zu den Testaufgaben. Wenn diese zu kompliziert oder nicht präzise genug formuliert werden, lösen die Lernenden sie unterschiedlich, weil ihnen nicht klar ist, was sie genau tun sollen. Aufgaben, die unterschiedlich gelöst werden, können natürlich nicht zuverlässig messen.

Die Reliabilität wird durch statistische Verfahren berechnet, nachdem der Test mit einer großen Anzahl von Lernenden auf der entsprechenden Leistungsstufe erprobt wurde. Bei formellen Prüfungen sind solche Erprobungen und die anschließenden statistischen Analysen unabdingbar. Bei Lernfortschrittstests im Unterricht können solche Erprobungen natürlich nicht durchgeführt werden. Deshalb sollte man bei Lernfortschrittstest zumindest sorgfältig darauf achten, dass die Arbeitsanweisungen klar formuliert sind und die Testaufgaben eindeutig zu lösen sind, um einigermaßen sicher zu sein, dass der Test die Leistungen zuverlässig misst.

1.3 Objektivität

Bei formellen Prüfungen betrifft dieses Kriterium die Testdurchführung und bedeutet, dass alle Kandidaten unter den gleichen, vorher festgelegten Bedingungen geprüft werden. Es muss z.B. genau festgelegt werden, wie viel Zeit zum Lösen der Aufgaben zur Verfügung steht, welche Punkte für die Testaufgaben vergeben werden und ob Hilfsmittel, wie zum Beispiel ein Lexikon, verwendet werden können.

Zum anderen betrifft dieses Kriterium die Bewertung der Leistungen; dies gilt für formelle Prüfungen und Lernfortschrittstests. Testaufgaben zum Lese- und Hörverstehen können objektiv bewertet werden, indem man geschlossene Aufgabenformate verwendet, wie zum Beispiel Multiple-Choice. Schreiben und Sprechen hingegen können nicht objektiv bewertet werden, aber Bewertungsanleitungen können die Subjektivität der Bewertung erheblich reduzieren.

2. Aufgabentypen

Bei den Aufgabentypen unterscheidet man zwischen geschlossenen, halb-offenen und offenen Aufgaben. Offene Aufgaben sind die Aufgaben zum Schreiben und Sprechen, in denen die Lernenden

zwar inhaltlich geleitet werden, aber frei sind in der Ausführung der Aufgabe. Bei halb-offenen Aufgaben müssen die Lernenden ein oder zwei Wörter, oder einen kurzen Satz schreiben. Diesen Aufgabentyp findet man häufig in Tests zum Hörverstehen. Und bei geschlossenen Aufgaben zur Überprüfung der rezeptiven Fertigkeiten müssen die Lernenden die richtige Lösung nur erkennen, aber selbst nichts schreiben.

Zu den geschlossenen Aufgabentypen gehören Multiple-Choice, Richtig / Falsch und Zuordnungsaufgaben. Im Folgenden soll verdeutlicht werden, welche Verstehensstrategien man mit welchem Aufgabentyp testen kann und was man beim Entwerfen der Items¹ beachten sollte.

2.1 Multiple-Choice

Multiple-Choice Items eignen sich zur Überprüfung des Global- und des Detailverstehens schriftlicher oder mündlicher Texte. Ein Multiple-Choice Item enthält drei bis vier Wahlmöglichkeiten, von denen eine die richtige Lösung ist; die anderen sind die Distraktoren, d.h. die falschen Lösungen.

Ob man drei oder vier Wahlmöglichkeiten anbietet, hängt von der Fertigkeit und vom Niveau des Tests ab. Beim Hörverstehen werden im Allgemeinen drei Wahlmöglichkeiten angeboten, da viergliedrige Items eine zu hohe Gedächtnisleistung erfordern würden. Beim Leseverstehen bietet man häufig dreigliedrige Items für die Stufen A1 und A2 des Referenzrahmens an und viergliedrige Items für die Stufen ab B1.

Bei Multiple-Choice sollten die Items in ihrer Reihenfolge dem Textinhalt folgen, d.h. die Items zu Abschnitt 1 des Textes kommen zuerst, dann die zu Abschnitt 2, usw. Wenn in einem Test zum Leseverstehen die Items nicht dem Textverlauf folgen, müssen die Lernenden unnötig suchen und verlieren dadurch wertvolle Zeit; und beim Hörverstehen können die Items nur gelöst werden, wenn sie dem Textverlauf folgen.

Wenn man das Detailverstehen mit Multiple-Choice Items überprüfen möchte, sollte man darauf achten, dass die Items sich auf wichtige Einzelinhalte des Textes beziehen und nicht auf irrelevante Details. Um dies sicherzustellen, sollten im Text die wichtigen Einzelinformationen unterstrichen werden und die Items dann zu diesen Textstellen geschrieben werden.

Außerdem sollte man bei **Multiple-Choice** Items Folgendes beachten:

- Es sollte nur eine richtige Lösung pro Item geben und die Lösung sollte eindeutig sein.
- Die Distraktoren sollten zwar eindeutig falsch sein, aber trotzdem noch so plausibel, dass sie von schwächeren Lernenden gewählt werden.
- Der einleitende Satz (Stamm vom engl. *stem*) kann entweder als Frage formuliert sein oder als unvollständiger Satz, der durch die Wahlmöglichkeiten ergänzt wird. Wenn die Wahlmöglichkeiten einen einleitenden Satz ergänzen, müssen die Ergänzungen mit dem Stamm einen grammatisch korrekten und sinnvollen Satz bilden.

¹ Als „Item“ bezeichnet man jede einzelne Testaufgabe, während mit „Aufgabe“ das Ganze bezeichnet wird, also zum Beispiel ein Lesetext mit 5 Multiple-Choice Items.

- Die Wahlmöglichkeiten sollen sich optisch ähnlich sehen. Es sollte vermieden werden, zwei oder drei von der Länge her ähnliche Wahlmöglichkeiten anzubieten und eine, die sich stark davon unterscheidet.
- Verneinungen sollten so weit wie möglich vermieden werden. Sollten sie sich nicht umgehen lassen, sollte man sie in den Stamm aufnehmen und fett drucken. Die Wahlmöglichkeiten sollten keine Verneinungen enthalten, da sie dort verwirren könnten oder übersehen werden könnten.
- Um die Leselast zu reduzieren, sollten Formulierungen, die sich in den Wahlmöglichkeiten wiederholen, dort herausgenommen und in den Stamm integriert werden.
- Gibt es mehrere Items zu einem Text, sollte jedes Item unabhängig sein von den anderen, d.h. die Lösung eines Items sollte nicht die Lösung der anderen Items beeinflussen.

2.2 Richtig/Falsch

Bei diesem Aufgabentyp muss der Lernende entscheiden, ob die Aussage auf den Textinhalt zutrifft oder nicht. Richtig/Falsch Items können sowohl das Globalverstehen als auch das Detailverstehen oder das selektive Verstehen überprüfen, je nachdem auf welche Textstellen das Item sich bezieht. Auch hier müssen die Items nach dem Text stehen, da die Lernenden zuerst den Text lesen bzw. hören und dann die Items lösen sollen. Außerdem müssen auch hier die Items in ihrer Reihenfolge dem Textinhalt folgen.

Als Aufgabe zum Leseverstehen eignet sich dieser Typ allerdings nur für Tests auf niedrigem sprachlichem Niveau, also A1 oder A2, da man in einem Text, den man lesen kann, die richtige Lösung leicht findet. Richtig/Falsch Items sind deswegen geeigneter zur Überprüfung des Hörverstehens auf der Grundstufe. Aber Richtig/Falsch Items haben den Nachteil, dass die Ratewahrscheinlichkeit hoch ist – sie liegt bei 50%. Man sollte deshalb in Tests nicht nur diesen Aufgabentyp verwenden.

Für die Stufen ab B2 eignet sich eine Variante dieses Aufgabentyps, nämlich „Ja / Nein / Text sagt dazu nichts“. Der *TestDaf* z.B. verwendet diesen Aufgabentyp in einer der Aufgaben zum Leseverstehen (<http://www.testdaf.de>). Dieser Aufgabentyp kann allerdings nur beim Leseverstehen verwendet werden, da man den Text sehen muss, um zu entscheiden, ob etwas im Text steht oder nicht. Und er ist auch nicht leicht zu entwerfen, denn die Items, d.h. die Aussagen zum Text, die sich auf „Nein“ oder „Text sagt dazu nichts“ beziehen, müssen eindeutig zu lösen sein. Ist die Lösung „Text sagt dazu nichts“, darf der Text also wirklich nichts dazu sagen, denn sonst wäre die Lösung „Nein“. Trotzdem müssen die Aussagen, die sich auf die Lösung „Text sagt dazu nichts“ beziehen, so plausibel sein, dass der Text etwas dazu sagen könnte.

2.3 Zuordnung

Zu diesem Aufgabentyp gehört eine Reihe von unterschiedlichen Aufgabenformen. Allen gemeinsam ist, dass die Lernenden zwei passende Teile einander zuordnen müssen. In ihrer einfachsten Form bestehen Zuordnungen darin, die Kenntnis der grammatischen Strukturen zu überprüfen, indem

Satzhälften aus einer Liste mit den richtigen Satzhälften aus einer anderen Liste verbunden werden müssen. Bei der Überprüfung des Lese- und Hörverstehens werden häufig Texte und Aussagen einander zugeordnet. Die Items können z.B. Aussagen über Personen und ihre Interessen sein und der Lernende muss für jede Person den passenden Text finden.

Als Textvorlage für Zuordnungsaufgaben dienen mehrere kürzere Texte. Mit dieser Aufgabe können unterschiedliche Verstehensstrategien getestet werden. Wenn ein Item so formuliert wird, dass es sich auf einen der Texte als Ganzes bezieht, testet man das Globalverstehen. Wird das Item hingegen so formuliert, dass der Lernende gezielt eine bestimmte Information in einem der Texte finden muss, testet man das selektive Verstehen. Beim Aufgabentyp „Überschriften kürzeren Texten zuordnen“ z.B. muss man die Texte nicht im Detail verstehen und man sucht auch nicht nach spezifischen Informationen. Dieser Aufgabentyp testet also das Globalverstehen. Beim Aufgabentyp „Personen mit bestimmten Interessen den passenden Texten zuordnen“ hingegen muss der Lernende die Texte zwar auch global verstehen, aber der Testfokus liegt auf dem Auffinden spezifischer Informationen.

Wenn man Zuordnungsaufgaben zum Leseverstehen entwickelt, stehen die Items vor den Texten. Die Lernenden lesen zuerst die Items und suchen dann in den Texten gezielt nach der Information, die sie benötigen, um das Item zu lösen. Und außerdem ist hier – im Gegensatz zu Multiple-Choice oder Richtig/Falsch - die Reihenfolge der Items beliebig.

Außerdem muss man bei diesem Aufgabentyp Folgendes beachten:

- die Zuordnung darf sich gegen Ende der Aufgabe nicht automatisch ergeben. Dies wäre zum Beispiel der Fall, wenn man sechs Personen nennt, denen sechs Texte zugeordnet werden sollen, denn es gäbe dann immer weniger Wahlmöglichkeiten und das letzte Item ergäbe sich automatisch (es wäre der Text, der übrig bleibt, vorausgesetzt alle anderen Items wurden richtig gelöst). Man muss also entweder mehr Personen definieren, von denen einige als Distraktoren dienen, oder man muss mehr Texte vorgeben als man Personen definiert; in diesem Fall wären einige der Texte die Distraktoren.

Diese drei Aufgabentypen: Multiple-Choice, Richtig-Falsch und Zuordnung eignen sich zum Testen des Lese- und Hörverstehens, weil sie die beiden Fertigkeiten weitestgehend isoliert von der Fertigkeit „Schreiben“ überprüfen.