



Erläuterungen zum Klassen- und Schulbericht im Enseignement fondamental und secondaire

1	Erklärung Kompetenzen	2
1.1	Was messen die ÉpStan? Was messen sie nicht?	2
1.2	Wie werden die Kompetenzen erfasst?	7
1.3	Was unterscheidet die ÉpStan von klassischen Prüfungen?	7
2	Wie zuverlässig messen die ÉpStan? Studienergebnisse aus dem Enseignement fondamental	8
3	Auswertung der Testergebnisse	9
3.1	ÉpStan-Metrik.....	9
3.2	Was ist der Erwartungsbereich?	10
4	Erklärung Lernmotivation	11
4.1	Lernmotivation im Fach	11
4.2	Generelle schulische Lernmotivation	12
4.3	Klassen- und Schulklima	12
5	Referenzen	13

1 Erklärung Kompetenzen

1.1 Was messen die ÉpStan? Was messen sie nicht?

Bei den ÉpStan werden die Leistungen der Schülerinnen und Schüler in ausgewählten Kompetenzbereichen anhand von standardisierten Tests erfasst. Die im Rahmen dieser Kompetenztests verwendeten Aufgaben basieren auf den Bildungsstandards, die vom MENJE für die verschiedenen Zyklen der Grund- und Sekundarschule, welche dem traditionellen nationalen Lehrplan folgen, entwickelt wurden. In diesen Bildungsstandards werden Kompetenzen folgendermaßen definiert: « *la capacité de mettre en œuvre un ensemble organisé de connaissances, d'habiletés et d'attitudes qu'un élève doit mobiliser dans un contexte donné lui permettant de fournir une réponse adéquate à une problématique complexe* » (Jonnaert zit. in MENFP, 2007, S. 3).

Folgt man dieser Definition, lassen sich die Kompetenzen der Schülerinnen und Schüler am besten evaluieren, indem sie Aufgaben mit komplexen Problemstellungen bearbeiten. Doch was genau bedeutet „komplex“ in diesem Zusammenhang? Die Lösung „komplexer Probleme“ erfordert den integrativen Einsatz von Wissen und Können, daher ist sie zumeist zeitintensiv, anstrengend und verlangt von den Schülerinnen und Schülern ein hohes Maß an Motivation.

Im Rahmen der ÉpStan unterscheiden wir verschiedene Grade an Komplexität, je nach Anforderungsprofil der Problemstellung. So kann das Ziel einer Leseaufgabe etwa darin bestehen, im Text explizierte Informationen zu lokalisieren und zu benennen (geringere Komplexität), oder aber aus dem Gelesenen eine Schlussfolgerung zu ziehen (höhere Komplexität).

Hervorzuheben ist an dieser Stelle auch, dass die ÉpStan-Kompetenztests nicht alle in den Bildungsstandards definierten Kompetenzen überprüfen. Denn zum einen lassen sich mithilfe von standardisierten Aufgabenformaten nicht alle Kompetenzbereiche erfassen. Zum anderen ist die Zeit, die zur Durchführung der ÉpStan zur Verfügung steht, sehr begrenzt. Aus diesen Gründen setzen sich die ÉpStan aus einer repräsentativen Auswahl von möglichst ökonomisch und standardisiert erfassbaren Schlüsselkompetenzen zusammen.

Natürlich werden anhand der ÉpStan-Aufgaben zum Lese- und ggf. Hörverstehen nur zwei Teilbereiche der sprachlichen Kompetenz der Schülerinnen und Schüler erfasst. Die beiden Kompetenzbereiche spielen im Schul- und Lernalltag der Schülerinnen und Schüler jedoch insofern eine zentrale Rolle, als erst die Fähigkeit zur Rezeption der Schriftsprache den Zugang zu weiterführender Bildung herstellt. Darüber hinaus lassen sich Lese- und ggf. Hörverstehen verhältnismäßig ökonomisch, in hohem Maße standardisiert und objektiv messen.

Erfasste Kompetenzen im Zyklus 2.1:

- **Mathematik im Zyklus 2.1.** Die Mathematikaufgaben beziehen sich inhaltlich auf (a) den Bereich „Raum und Form“ (MENFP, 2011, S. 26, 116), (b) den Bereich „Zahlen und Operationen“ (MENFP, 2011, S. 28, 117), sowie (c) den Bereich „Größen und Maße“ (MENFP, 2011, S. 30, 118).

- **Erste Schritte zur Schriftsprache im Zyklus 2.1.** Die Testaufgaben zur Erfassung dieser Vorläuferfertigkeiten (MENFP, 2011, S. 8) beziehen sich hauptsächlich auf folgende Kompetenzen: Phonologische Bewusstheit, auditive Aufmerksamkeit, visuelle Diskrimination, und erstes Lesen und Schreiben (z.B. Buchstaben und einfache ein- und zweisilbige Wörter).
- **Luxemburgisch-Hörverstehen im Zyklus 2.1.** Die Testaufgaben zur Erfassung des Hörverstehens im Luxemburgischen (MENFP, 2011, S. 6) beziehen sich auf verschiedene Typen von Hörtexten (z. B. Dialog, Erzählung, Reportage). Es werden Aufgaben zu verschiedenen Kompetenzen entwickelt (vgl. MENFP, 2011, S. 6), z. B. seinen Gesprächspartner verstehen, Informationen lokalisieren und verstehen, sowie Hörstrategien anwenden (z.B. Geräusche erkennen).
- **Deutsch-Hörverstehen im Zyklus 2.1.** Im Test zur Erfassung des Deutsch-Hörverstehens im Zyklus 2 (MENFP, 2011, S. 6) werden die Schülerinnen und Schüler mit verschiedenen Hörtext-Formaten konfrontiert (z. B. mit Gesprächen, kurzen Erzählungen oder Kindersendungen). Zu diesen Hörtexten werden ihnen Fragen gestellt, die auf verschiedene Fertigkeiten abzielen (vgl. MENFP, 2011, S. 12): z. B. darauf, Gesprächen zu folgen sowie einen Gesprächspartner zu verstehen (der Anweisungen gibt), kurze Hörtexte global zu verstehen (wie etwa in dieser Frage: Was machen die Kinder in der Geschichte?), Einzelinformationen zu lokalisieren und verstehen (wie in der Frage: Wo treffen sich die Kinder?) sowie Hörstrategien anzuwenden (indem z.B. Geräusche erkannt werden).

Erfasste Kompetenzen im Zyklus 3.1:

- **Mathematik im Zyklus 3.1.** Die Mathematikaufgaben beziehen sich inhaltlich auf (a) den Bereich „Zahlen und Operationen“ (MENFP, 2011, S. 28-29, 121) sowie (b) einen kombinierten Bereich aus „Raum und Form“ (MENFP, 2011, S. 26-27, 120) und „Größen und Maße“ (MENFP, 2011, S. 30-31, 122). Weiterhin werden zwei Prozesskompetenzen unterschieden, die auf (a) „Problemlösen und Modellieren“ (vgl. MENFP, 2011, S. 115) sowie (b) „spezifische Grundfertigkeiten“ (vgl. MENFP, 2008, S. 11) abzielen. Letztere sind definiert als mathematische Kenntnisse und Fertigkeiten, die isoliert, kontextfrei und ohne jegliche Transferarbeit angewendet werden. Durch die Unterscheidung zwischen kontextualisierten Aufgaben (Problemlösen und Modellieren) und dekontextualisierten Aufgaben (spezifische Grundfertigkeiten) wird der Inhaltsbereich „Sachrechnen“ (MENFP, 2011, S. 32-33, 123) implizit mit abgedeckt.
- **Deutsch-Leseverstehen im Zyklus 3.1.** Die Testaufgaben zur Erfassung der Deutsch-Lesekompetenz (MENFP, 2011, S. 16-17) beziehen sich auf zwei verschiedene Textsorten: (a) Kontinuierliche Texte (z. B. Erzählung, Märchen, Sachtext) und (b) diskontinuierliche Texte (z. B. Rezept, Anleitung) sind (ähnlich wie bei PISA) im Verhältnis zwei zu eins vertreten. In der Testentwicklung werden weiterhin zwei Unterkompetenzen unterschieden, die auf (a) „Informationen lokalisieren und verstehen“ (vgl. MENFP, 2011, S. 82-83) und (b) „Informationen interpretieren und Lesestrategien kennen und nutzen“ (vgl. MENFP, 2011 S. 82-83) abzielen.

- **Deutsch-Hörverstehen im Zyklus 3.1.** Die Testaufgaben zur Erfassung der Deutsch-Hörkompetenz (MENFP, 2011, S. 12-13) beziehen sich auf verschiedene Typen von Hörtexten (z. B. Dialog, Reportage, Hörspiel). In der Testentwicklung werden zwei Unterkompetenzen unterschieden, die auf (a) „Informationen lokalisieren und verstehen“ (vgl. MENFP, 2011, S. 79-80) und (b) „Informationen interpretieren und Hörstrategien kennen und nutzen“ (vgl. MENFP, 2011 S. 79-80) abzielen.

Erfasste Kompetenzen im Zyklus 4.1:

- **Mathematik im Zyklus 4.1.** Die Mathematikaufgaben beziehen sich inhaltlich auf (a) den Bereich „Zahlen und Operationen“ (MENFP, 2011, S. 28-29, 125) sowie (b) einen kombinierten Bereich aus „Raum und Form“ (MENFP, 2011, S. 26-27, 124) und „Größen und Maße“ (MENFP, 2011, S. 30-31, 126). Weiterhin werden zwei Prozesskompetenzen unterschieden, die auf (a) „Problemlösen und Modellieren“ (vgl. MENFP, 2011, S. 115) sowie (b) „spezifische Grundfertigkeiten“ (vgl. MENFP, 2008, S. 11) abzielen. Letztere sind definiert als mathematische Kenntnisse und Fertigkeiten, die isoliert, kontextfrei und ohne jegliche Transferarbeit angewendet werden. Durch die Unterscheidung zwischen kontextualisierten Aufgaben (Problemlösen und Modellieren) und dekontextualisierten Aufgaben (spezifische Grundfertigkeiten) wird der Inhaltsbereich „Sachrechnen“ (MENFP, 2011, S. 32-33, 127) implizit mit abgedeckt.
- **Deutsch-Leseverstehen im Zyklus 4.1.** Die Testaufgaben zur Erfassung der Deutsch-Lesekompetenz (MENFP, 2011, S. 16-17) beziehen sich auf zwei verschiedene Textsorten: (a) Kontinuierliche Texte (z. B. Erzählung, Gedicht, Sachtext) und (b) diskontinuierliche Texte (z. B. Reportage, Kommentar). Diese sind (ähnlich wie bei PISA) im Verhältnis zwei zu eins vertreten. In der Testentwicklung werden weiterhin zwei Unterkompetenzen unterschieden, die auf (a) „Informationen lokalisieren und verstehen“ (vgl. MENFP, 2011, S. 88-89) und (b) „Informationen interpretieren und Lesestrategien kennen und nutzen“ (vgl. MENFP, 2011 S. 88-89) abzielen.
- **Französisch-Leseverstehen im Zyklus 4.1.** Die Testaufgaben zur Erfassung der Französisch-Lesekompetenz (MENFP, 2011, S. 24-25) beziehen sich auf zwei verschiedene Textsorten: (a) Kontinuierliche Texte (z. B. Erzählung, Märchen) und (b) diskontinuierliche Texte (z. B. Rezept, Anzeige). Diese sind (ähnlich wie bei PISA) im Verhältnis zwei zu eins vertreten. In der Testentwicklung werden weiterhin zwei Unterkompetenzen unterschieden, die auf (a) „Informationen lokalisieren und verstehen“ (vgl. MENFP, 2011, S. 104-105) und (b) „Informationen interpretieren und Lesestrategien kennen und nutzen“ (vgl. MENFP, 2011 S. 104-105) abzielen.

Erfasste Kompetenzen in der 7^e:

- **Mathematik in der 7^e.** Der Mathematiktest in der 7^e deckt die am Ende des vierten Zyklus der Grundschule erwarteten Kompetenzen in den vier Teilbereichen „Zahlen und Operationen“ (MENFP, 2011, S. 28-29, 129), „Raum und Form“ (MENFP, 2011, S. 26-27, 128), „Größen und Maße“ (MENFP, 2011, S.

30-31, 130) und „Sachrechnen“ (MENFP, 2011, S. 32-33, 131) ab. Der Test besteht aus 23 bis 28 Kurztextproblemen, die auf Französisch oder Deutsch präsentiert werden (die Schüler können jederzeit die Sprache wechseln und so stets die Sprache wählen, die ihnen zur Problemlösung am besten geeignet erscheint). Bei den meisten Aufgaben handelt es sich um halboffene oder Multiple-Choice Aufgaben. Sie sollen zeigen, inwieweit die Schüler in der Lage sind, die gelehrt Inhalte in bekannten und neuen mathematischen Situationen korrekt anzuwenden.

- **Deutsch-Leseverstehen in der 7^e.** Der Test zur Erfassung der Deutsch-Lesekompetenz umfasst die beiden Teilfertigkeiten „Informationen in einem Text finden und nutzen“ (Kompetenz 1 – K1) sowie „Informationen interpretieren und Lesestrategien / -techniken mobilisieren“ (Kompetenz 2 – K2) (MENFP, 2011, S. 16-17). Der Test besteht aus insgesamt vier bis sechs fortlaufenden (z. B. Erzählungen oder Sachtexte) oder diskontinuierlichen (z. B. Tabellen, Illustrationen oder Kartenmaterial) Texten, mit einer Länge von jeweils etwa 100 bis 550 Wörtern, die Themen aus dem Lebensalltag der Schülerinnen und Schüler wie Hobbies, Gesellschaft, Freundschaft oder Familie aufgreifen. Die Schwierigkeit des Tests orientiert sich an den Vorgaben für den vorangegangenen Lernzyklus der Grundschule (Zyklus 4).
- **Französisch-Leseverstehen in der 7^e.** Der Test zur Erfassung der Französisch-Lesekompetenz in der 7^e basiert auf den zwei Teilfertigkeiten „Informationen in einem Text finden und nutzen“ (Kompetenz 1 – K1) sowie „Informationen interpretieren/analysieren und Lesestrategien /-techniken mobilisieren“ (Kompetenz 2- K2) (s. Kompetenzraster und Entwicklungsstufen, Grundschule, Zyklen 1 bis 4, MENFP, 2011, S. 22-23). Der Test besteht aus insgesamt fünf bis sieben fortlaufenden (z.B. einfache Geschichten, Erzählungen oder Sachtexte) oder diskontinuierlichen (z.B. Illustrationen, Anzeigen) Texten, mit einer Länge von jeweils 50 bis 500 Wörtern. Die Texte behandeln alltagsnahe Themen wie Familie, Schule oder Freundschaft. Die Schwierigkeit des Tests orientiert sich an den Vorgaben für den vorangegangenen Lernzyklus der Grundschule (Zyklus 4).

Die Testdauer beträgt für alle Schülerinnen und Schüler 45 Minuten pro Fach, wobei die Reihenfolge der Tests variiert.

Erfasste Kompetenzen in der 5e:

- **Mathematik in der 5^e.** Der Test zur Erfassung der mathematischen Kompetenzen fokussiert auf Aufgaben zu mathematischen Modellen und Problemen, die sich deutlich in ihrem Komplexitätsgrad unterscheiden können. Alle eingesetzten Aufgaben lassen sich einem von drei mathematischen Inhaltsbereichen zuordnen: (1) Zahlen und Operationen (MENFP 2008b; 2008c, S. 21-23), (2) Ebene und räumliche Figuren (MENFP 2008b; 2008c, S. 17-19) oder (3) Abhängigkeit und Veränderung (MENFP 2008b; 2008c, S. 25-27). Die Schülerinnen und Schüler können die Mathematikaufgaben jeweils auf Deutsch oder Französisch bearbeiten.
- **Deutsch-Leseverstehen in der 5^e.** Der Test zur Erfassung der Deutsch-Lesekompetenz (MENFP 2008a, S. 34-35, 39-41, 50-52, 68) umfasst die beiden

Teilfertigkeiten „Informationen in einem Text finden und verstehen“ (Kompetenz 1 – K1) sowie „Texte analysieren, interpretieren und (wissensbasierte) Schlussfolgerungen ziehen“ (Kompetenz 2 – K2). Diese Teilkompetenzen lassen sich aus der Beschreibung der vier Lernstufen in den Bildungsstandards ableiten (vgl. MENFP 2008a, S. 68).

Jeder Deutsch-Test besteht aus insgesamt fünf bis acht fortlaufenden (z. B. Erzählungen oder Sachtexte) oder diskontinuierlichen Texten (z. B. Tabellen, Illustrationen oder Bauanleitungen) mit einer Länge von jeweils etwa 100 bis 750 Wörtern, die Themen wie Hobbies, Gesellschaft, Freundschaft oder Berufsausbildung behandeln.

- **Französisch-Leseverstehen in der 5^e.** Der Test zur Erfassung der Französisch-Lesekompetenz basiert ebenfalls auf den zwei Teilkompetenzen „Informationen in einem Text finden und verstehen“ (Kompetenz 1 – K1) sowie „Texte analysieren, interpretieren und (wissensbasierte) Schlussfolgerungen ziehen“ (Kompetenz 2 – K2). Denn auch für Französisch lassen sich diese Teilkompetenzen aus der Beschreibung der vier Lernstufen in den Bildungsstandards ableiten (vgl. MENFP 2008a, S. 68). Jeder Französisch-Test besteht aus insgesamt vier bis sieben fortlaufenden (z.B. Geschichten oder Sachtexte) oder diskontinuierlichen Texten (z.B. Tabellen oder Illustrationen) mit einer Länge von jeweils etwa 100 bis 750 Wörtern, die Themen wie Freundschaft, Hobby, Familie oder Berufsausbildung behandeln.

Für die drei Schulformen „Enseignement secondaire classique“, „Enseignement secondaire général – voie d’orientation“ sowie „Enseignement secondaire général – voie de préparation“ werden unterschiedliche Testversionen erstellt, die stets eine Mindestanzahl an sich überschneidendem Text- und Aufgabenmaterial enthalten, um ihre Vergleichbarkeit zu gewährleisten.

Dementsprechend werden für alle drei Kompetenzbereiche jeweils drei Testversionen erstellt. Die Testversionen heißen den jeweiligen Schulformen entsprechend:

- *Enseignement secondaire classique (ESC)*
- *Enseignement secondaire général – voie d’orientation (ESG)*
- *Enseignement secondaire général – voie de préparation (ESG-VP)*

Jede Testversion besteht einerseits aus Aufgaben, die speziell auf das durchschnittliche Leistungsniveau der jeweiligen Schulform zugeschnitten sind. Andererseits enthält jede Testversion auch immer eine Mindestanzahl an Aufgaben, die von allen Schülerinnen und Schülern bearbeitet werden. Auf diese Weise können mithilfe moderner psychometrischer Verfahren schulformübergreifende Kompetenzvergleiche angestellt werden.

Unabhängig von der Testversion beträgt die Testzeit für alle Schülerinnen und Schüler 45 Minuten pro Fach, wobei die Reihenfolge der Tests zufallsbasiert variiert.

Die ÉpStan liefern somit eine hochstandardisierte, aber unvollständige Momentaufnahme, während die *Bilans* bzw. Zeugnisnoten auf einer langfristigen und vollständigen Evaluation der schulischen Leistungen beruhen. Dementsprechend sind die Ergebnisse aus den ÉpStan komplementär und nicht in Konkurrenz zu den *Bilans* bzw. Zeugnisnoten zu verstehen.

1.2 Wie werden die Kompetenzen erfasst?

In Bezug auf Schulen, die dem traditionellen Lehrplan folgen, ermöglichen die ÉpStan es zu erfassen, ob die Schülerinnen und Schüler zu Beginn eines jeden Zyklus die im traditionellen nationalen Lehrplan definierten Socles des vorherigen Zyklus erreicht haben. Für Internationale Öffentliche Schulen, die einem alternativen Lehrplan folgen (z.B. dem Lehrplan des europäischen Schulsystems), sollten diese definierten Socles dementsprechend nur als breiter Referenzrahmen betrachtet werden. Bitte bedenken Sie bei der Interpretation von Grafiken (z.B. Kompetenzverteilung der Schülerinnen und Schüler in Ihrer Klasse), dass von Ihren Schülerinnen und Schülern nicht erwartet werden kann, alle im traditionellen nationalen Lehrplan unterrichteten Kompetenzen erworben zu haben. So ist es möglich, dass ihre tatsächlichen Kompetenzen in einigen Bereichen über den nationalen Referenzrahmen (Socles) hinausgehen, während sie in anderen dahinter zurückbleiben.

In einer ersten Phase wird jede Aufgabe einer theoretischen Schwierigkeitsstufe (z.B. „unter Niveau Socle“, „Niveau Socle“, „Niveau Avancé“, „über Niveau Avancé“) zugeordnet. Diese Erstklassifizierung der Aufgaben erfolgt in den Testentwicklungsgruppen unter strikter Berücksichtigung der jeweiligen Deskriptoren (MENFP, 2011 & MENFP 2008a; 2008b; 2008c) und noch bevor empirische Schwierigkeitsindikatoren für die Aufgaben vorliegen. In einer zweiten Phase werden nach der Datenerhebung die theoretischen und empirischen Aufgabenschwierigkeiten miteinander verglichen. Wenn diese für eine Aufgabe nicht übereinstimmen, wird entweder eine theoretische Reklassifikation vorgenommen – sofern dies anhand der theoretischen Deskriptoren zu den einzelnen Kompetenzbereichen und -stufen vertretbar scheint – oder die besagte Aufgabe wird bei der endgültigen Berechnung des Kompetenzniveaus nicht berücksichtigt.

Durch diese hier stark vereinfacht dargestellte Prozedur können die Testentwicklungsgruppen sicherstellen, dass theoretisch schwierige bzw. einfache Aufgaben auch empirisch schwierig bzw. einfach ausfallen. Außerdem wird so dafür gesorgt, dass die empirischen Schwierigkeiten der jeweiligen Niveau- bzw. Lernstufen relativ ähnlich ausfallen, sodass sich z.B. alle Aufgaben, die zur Berechnung des *Niveau Socle* oder der Lernstufe 2 in Mathematik herangezogen werden, in etwa im gleichen empirischen Schwierigkeitsbereich bewegen.

1.3 Was unterscheidet die ÉpStan von klassischen Prüfungen?

In klassischen Prüfungen wird in erster Linie der in der Klasse gelehrt Stoff abgefragt, den die Schülerinnen und Schüler beherrschen sollten. Bei diesem Prüfungsformat wird (prinzipiell) davon ausgegangen, dass Schülerinnen und Schüler, die am Unterricht teilgenommen haben, alle Aufgaben lösen können. Wurden alle Aufgaben richtig gelöst, erfüllen die Schülerinnen und Schüler die Grunderwartung, den Sollzustand. In der Regel gehen klassische Prüfungen jedoch nicht über dieses zu überprüfende Soll

hinaus, sodass sie sehr guten Schülerinnen und Schüler kaum Möglichkeiten bieten, ihr Wissen und Können unter Beweis zu stellen. Nur wenn die Leistungen der Prüflinge *nicht* den angestrebten Sollzustand erreichen, wird dies durch klassische Prüfungsformen ersichtlich. Diese defizitorientierte Herangehensweise ist sehr stark in unserer Prüfungskultur verankert. So wird bei der Verbesserung von Prüfungen kaum darauf aufmerksam gemacht, was die Schülerinnen und Schüler richtig machen. Die Korrektur hebt vor allem hervor, was ihnen (noch) nicht gelingt. Im Allgemeinen kommt es bei klassischen Prüfungen also eher weniger darauf an, was Schülerinnen und Schüler können, als darauf, ihre Schwächen bzw. ihre Abweichung vom Soll zu identifizieren.

Im Gegensatz dazu decken die ÉpStan die komplette Bandbreite an unterschiedlich schwierigen Aufgaben ab: Das, was fast jeder kann, das was jeder (mindestens) können sollte und das, was (weit) darüber hinausgeht. Dementsprechend ist von vornherein nicht davon auszugehen, dass alle Schülerinnen und Schüler alle ÉpStan-Aufgaben richtig lösen. Vielmehr sollen sie ihre Kompetenzen voll ausschöpfen, indem sie möglichst viele (unterschiedlich schwierige) Aufgaben bearbeiten. Bei den ÉpStan kommt es also nicht vorrangig darauf an, Schwächen zu identifizieren – in erster Linie soll festgestellt werden, über welche (Teil-) Kompetenzen die Schülerinnen und Schüler bereits verfügen.

Insofern beruht der bei den ÉpStan verfolgte Prüfungsansatz eindeutig auf den Prinzipien des kompetenzorientierten Unterrichts. Das bedeutet jedoch nicht, dass die eingesetzten Aufgaben pauschal als repräsentativ für diese Form des Unterrichts gelten können. Der Sachverhalt ist etwas komplizierter.

Die ÉpStan zielen auf die Vergleichbarkeit der Schülerleistungen untereinander sowie mit einem Standard ab. Vor diesem Hintergrund wurden die getesteten Kompetenzbereiche ausgewählt, weil sie a) von zentraler Bedeutung für den Schul- und Lernalltag sind und weil sie sich b) mit verhältnismäßig wenig Aufwand standardisiert, objektiv und zuverlässig erfassen lassen. Das Gleiche gilt für die ausgewählten Aufgabenformate. Daher werden bei den ÉpStan überwiegend geschlossene und halboffene Aufgaben eingesetzt. In der Unterrichtssituation steht hingegen der Lernprozess der Schülerinnen und Schüler im Mittelpunkt, und dieser ist vorzugsweise mit offenen Aufgabenformaten zu fördern.

2 Wie zuverlässig messen die ÉpStan? Studienergebnisse aus dem Enseignement fondamental

Ein zentrales Ziel der ÉpStan ist, die Schülerkompetenzen so standardisiert wie möglich, d.h. objektiv und zuverlässig – unter Berücksichtigung der gegebenen Rahmenbedingungen – zu erfassen. Maßnahmen, wie die ausschließliche Anwendung von geschlossenen und halboffenen Aufgabenformaten, die Bereitstellung detaillierter Korrekturanweisungen (zur Vermeidung einer zu milden bzw. zu strengen Korrektur) sowie einer digitalen Korrekturmaske, sollen sicherstellen, dass dieses Ziel erreicht wird. Es stellt sich natürlich die Frage, ob unterschiedliche Korrektoren zu den gleichen Kompetenzeinschätzungen für die einzelnen Schülerinnen und Schüler kommen.

Um dieser Frage nachzugehen, haben wir bei den ÉpStan 2010/11 (für 3.1) sowie bei den ÉpStan 2015/16 (für 2.1 und 3.1) eine Zusatzstudie durchgeführt. Pro Jahrgang

wurde von jedem der vier Testhefte eine zufällige Stichprobe von 30 bearbeiteten Schülerheften gezogen. In einer ersten Phase wurden in diesen Heften alle vom Klassenlehrer eingetragenen Korrekturen unkenntlich gemacht. In einer zweiten Phase wurden die jeweils 120 Schülerhefte von einem zweiten, hierfür geschulten Korrektor streng nach den im Manual vorgegebenen Korrekturanweisungen erneut verbessert. Insgesamt lagen so für 3400 (ÉpStan 2010/11) bzw. 3000 (ÉpStan 2015/16, Zyklus 2.1) und 4400 (ÉpStan 2015/16, Zyklus 3.1) Schülerantworten jeweils Verbesserungen von zwei Korrektoren vor. Das wichtigste Ergebnis war, dass die Korrekturen in höchstem Maße übereinstimmten und diese Übereinstimmung über die zwei Jahrgänge hinweg stabil blieb. Absolut lag die Übereinstimmung bei 97,4% (ÉpStan 2010/11) bzw. bei 97,2% (ÉpStan 2015/16, Zyklus 2.1) und 98,0% (ÉpStan 2015/16, Zyklus 3.1); berücksichtigt man, dass die Übereinstimmung zwischen den Korrektoren auch zufallsbedingt sein könnte, lag die korrigierte Übereinstimmung immer noch bei 95,3% (ÉpStan 2010/11) bzw. bei 96,1% (ÉpStan 2015/16, Zyklus 2.1) und 94,9% (ÉpStan 2015/16, Zyklus 3.1). Zusammenfassend zeigt sich also, dass die präzisen Korrekturanweisungen und vor allem deren sorgfältige Anwendung durch die Lehrkräfte es ermöglichen, die Schülerkompetenzen äußerst zuverlässig und objektiv zu erfassen. Da die Prozeduren der ÉpStan im Zyklus 4.1 quasi eins-zu-eins aus dem Zyklus 3.1 übernommen wurden, können wir mit hoher Wahrscheinlichkeit davon ausgehen, dass der positive Befund aus dem Zyklus 3.1 auch auf den Zyklus 4.1 zutrifft. Im Enseignement secondaire werden die Antworten der Schülerinnen und Schüler direkt in der Testplattform gespeichert und weiterverarbeitet. Es ist also kein händisches Kodieren seitens der Lehrkräfte notwendig.

3 Auswertung der Testergebnisse

3.1 ÉpStan-Metrik

Das übergreifende Ziel der ÉpStan besteht darin, die Leistungen der Schülerinnen und Schüler längsschnittlich – das heißt, über die Zeit hinweg - zu vergleichen, um mögliche Effekte von Reformen im Bildungswesen sichtbar zu machen. Hierfür können jedoch nicht in jedem Schuljahr die gleichen Tests eingesetzt werden, da ansonsten die Gefahr bestünde, dass Schülerinnen und Schüler gezielt auf den Test vorbereitet und die Testergebnisse verzerrt würden. Aus diesem Grund werden die Leistungen der Schülerinnen und Schüler bei den ÉpStan nicht einfach zu Punktsummen zusammengefasst, sondern Kompetenzwerte ermittelt, die mit statistischen Modellen aus der sogenannten *Item Response Theory* (IRT) berechnet werden. Diese Kompetenzwerte können über verschiedene Schülerjahrgänge und ggf. Schulformen hinweg verglichen werden, auch wenn in den jeweiligen Kompetenztests Jahr für Jahr andere Aufgaben vorkommen. Voraussetzung hierfür ist lediglich, dass einige Aufgaben bereits in früheren Testversionen eingesetzt wurden.

Die Kompetenzwerte wurden so normiert, dass der Mittelwert aller luxemburgischen Schülerinnen und Schüler (schulformübergreifend) bei 500 Punkten liegt. Die Standardabweichung (also die mittlere Abweichung der Testwerte vom Mittelwert) wurde auf 100 Punkte gesetzt. Je nach Lernzyklus, gibt es ein anderes Referenzjahr, welches davon abhängig ist, seit wann die einzelnen Klassenstufen bei den ÉpStan teilnehmen: Für den Zyklus 2.1 ist es das Schuljahr 2014/15, für den Zyklus 3.1 das Schuljahr 2011/12, für den Zyklus 4.1 das Schuljahr 2017/18 in den der Sekundarschule für die 7e

das Schuljahr 2018/19¹ sowie für die 5e das Schuljahr 2011/2012. Dieses Schuljahr dient demnach als Ausgangspunkt und erster Referenzjahr, bis zu dem sich die diesjährigen ÉpStan-Ergebnisse vergleichen lassen. Wenn sich also der mittlere ÉpStan-Wert einer bestimmten Gruppe von Schülerinnen und Schülern von einem auf das andere Jahr verändern, deutet dies auf eine tatsächliche Veränderung der Kompetenzen in dieser Gruppe hin.

3.2 Was ist der Erwartungsbereich?

Viele Lehrkräfte wissen von vornherein, dass die Leistungen ihrer Klasse aufgrund der soziokulturellen und sozioökonomischen Zusammensetzung ziemlich sicher unter oder über dem Landesdurchschnitt liegen werden. Aber wozu werden die Leistungen überprüft, wenn die Ergebnisse sowieso schon festzustehen scheinen?

Zahlreiche internationale und nationale Studien (z. B. PISA, PIRLS, die ÉpStan) haben gezeigt, dass Merkmale wie Geschlecht, sozialer Hintergrund und Muttersprache deutlich mit den Leistungen in Kompetenztests zusammenhängen. Eine höhere durchschnittliche Leistung im Deutsch-Leseverstehen erzielen in Luxemburg zum Beispiel Jugendliche mit luxemburgischer oder deutscher Muttersprache, einem hohen sozialen Status und/oder weiblichem Geschlecht. Auf der anderen Seite erbringen Jugendliche mit portugiesischer oder französischer Muttersprache, einem niedrigen sozialen Status und/oder männlichem Geschlecht in diesem Kompetenzbereich im Mittel niedrigere Leistungen. Die soziokulturelle und sozioökonomische Zusammensetzung der Schülerschaft in Luxemburg unterscheidet sich jedoch erheblich zwischen einzelnen Schulen und Klassen, sodass sich die Leistungen einzelner Klassen nicht ohne Weiteres vergleichen lassen.

Am einfachsten lässt sich dies wohl an einem (stark vereinfachten) Beispiel erläutern: Stellen wir uns eine fiktive Schulklasse vor, die einen Deutschttest absolviert hat. Die Schülerinnen und Schüler haben in diesem Test durchschnittlich 430 Punkte erreicht (die Punktzahlen beziehen sich auf die ÉpStan-Metrik aus Abschnitt 3.1) und liegen damit deutlich unter dem nationalen Durchschnitt. Nehmen wir weiter an, dass die Klasse vor allem von Jungen aus sozial benachteiligten Elternhäusern besucht wird, die zu Hause Portugiesisch oder Französisch sprechen. Vor diesem Hintergrund ist das unterdurchschnittliche Abschneiden der Klasse nicht überraschend. Die interessantere Frage lautet daher: Wie verhält sich das tatsächliche Ergebnis der Klasse zu demjenigen, das aufgrund der Zusammensetzung der Schülerschaft zu erwarten gewesen wäre? Um diese Frage zu beantworten, berechnen wir „Erwartungsbereiche“ anhand eines statistischen Modells (dem sogenannten Regressionsmodell).

Dabei gehen wir folgendermaßen vor: Zunächst ermitteln wir den Zusammenhang zwischen den Testleistungen der Schülerinnen und Schüler in den gemessenen Kompetenzbereichen einerseits und den Merkmalen Geschlecht, sozioökonomischer Hintergrund, Muttersprache, Migrationshintergrund, Schulform (in der Sekundarstufe) und Schullaufbahn Daten andererseits. Die genannten Merkmale bezeichnen wir als Hintergrundvariablen. Für jede dieser Hintergrundvariablen liefert das

¹ Es ist allerdings zu beachten, dass nur etwa zwei Drittel aller Schülerinnen und Schüler der 7e an den ÉpStan teilgenommen haben. Bei der Interpretation der Ergebnisse ist deshalb zu berücksichtigen, dass die getesteten SchülerInnen und Schüler nicht den ganzen 7e-Jahrgang repräsentieren und sich die Grenze zur Festlegung des Niveau *Socle* in den kommenden Jahren noch verändern könnte.

Regressionsmodell eine Gewichtung, genannt Parameter, welche angibt, wie stark dieses Merkmal die Testleistung beeinflusst.

Im zweiten Schritt verwenden wir diese Parameter, um für jede Klasse einen Erwartungsbereich zu berechnen, in dem das Ergebnis der Klasse liegen sollte, sofern die im Regressionsmodell abgebildeten Zusammenhänge die tatsächliche Zusammensetzung der Schülerschaft in Luxemburg widerspiegeln. Die Breite dieses Erwartungsbereichs spiegelt die Messgenauigkeit der eingesetzten Tests, die Vorhersagegenauigkeit des Regressionsmodells sowie die Unsicherheit aufgrund fehlender Daten wider. Hat eine Klasse nun ein Ergebnis erzielt, das über dem Erwartungsbereich liegt, so hat sie damit besser abgeschnitten, als aufgrund der sozioökonomischen und soziokulturellen Merkmale der Schülerinnen und Schüler statistisch zu erwarten gewesen wäre. Ein Ergebnis unterhalb des Erwartungsbereichs signalisiert hingegen ein schlechteres Resultat, als statistisch zu erwarten gewesen wäre.

Bei den ÉpStan wird also das Leistungsprofil jeder einzelnen Klasse in Bezug zu Klassen mit vergleichbarer soziokultureller und sozioökonomischer Zusammensetzung gesetzt. Auf diese Weise erhalten z.B. auch Lehrkräfte, die sehr effektiv unter „schwierigen“ Bedingungen arbeiten, eine positive und wissenschaftlich abgesicherte Rückmeldung zum Leistungsstand ihrer Klasse. Grundsätzlich sind Erwartungsbereiche jedoch für *alle Lehrkräfte* interessant, die sich ein faires Bild von den Kompetenzen ihrer Schülerinnen und Schüler machen möchten.

4 Erklärung Lernmotivation

Zusätzlich zu den Leistungstests bearbeiten die Schülerinnen und Schüler einen Fragebogen. Dieser erfragt folgende Bereiche:

- Lernmotivation in den bei den ÉpStan abgeprüften Kompetenzbereichen
- generelle Lernmotivation in der Schule
- Klassen- und Schulklima

Die Lernmotivation der Schülerinnen und Schüler wird anhand von Aussagen wie „Ich interessiere mich für die meisten Schulfächer“ erfasst. Durch Ankreuzen geben die Schülerinnen und Schüler den Grad ihrer Zustimmung zu diesen Aussagen an. Hierfür steht ihnen eine vierstufige Antwortskala mit Antwortkategorien von „trifft nicht zu“ bis „trifft zu“ zur Verfügung. Die Antworten der verschiedenen Aussagen werden dann thematisch zu Skalenwerten zusammengefasst (z. B. in einer Skala für generelles Schulinteresse). In der Sekundarstufe können die Schülerinnen und Schüler den Fragebogen wahlweise auf Deutsch oder Französisch bearbeiten.

4.1 Lernmotivation im Fach

Die Lernmotivation im Fach wird anhand von vier Skalen erfasst: Selbstkonzept, Interesse, Angst und Leseinteresse (letzteres nur für die Sprachfächer).

- Die Skala Selbstkonzept gibt Aufschluss über die Einschätzung der Schülerinnen und Schüler bezüglich ihrer eigenen schulischen Leistungen und Fähigkeiten im jeweiligen Fach. Zur Messung des fachspezifischen Selbstkonzepts nehmen die Schülerinnen und Schüler Stellung zu Aussagen wie: „Ich bin gut im Fach Rechnen/Mathematik.“
- Die Skala Interesse gibt an, inwiefern sich die Schülerinnen und Schüler für dieses Fach interessieren und sie daran Freude haben. Eine Beispielaussage hierfür lautet: „Ich interessiere mich für das Fach Rechnen/Mathematik.“
- Die Skala Angst informiert darüber, inwiefern dieses Fach den Schülerinnen und Schülern Angst macht, Sorgen bereitet oder inwiefern Prüfungen in diesem Fach sie nervös machen. Eine Beispielaussage hierfür ist: „Ich habe Angst vor dem Fach Rechnen/Mathematik.“
- Die Skala Leseinteresse gibt an, inwiefern die Schülerinnen und Schüler gerne lesen. Eine Beispielaussage hierfür ist: „Ich lese gerne auf Deutsch.“

Die Mittelwerte sind wie folgt zu lesen: Je höher sie sind, umso positiver ist das Selbstkonzept der Schülerinnen und Schüler, ihr Interesse am Fach oder ihr Leseinteresse – aber auch ihre Angst in einem bestimmten Schulfach.

4.2 Generelle schulische Lernmotivation

Die generelle schulische Lernmotivation der Schülerinnen und Schüler wird anhand von drei Skalen erfasst:

- Selbstkonzept (z. B. „Ich bin gut in den meisten Schulfächern.“)
- Interesse (z. B. „Die meisten Schulfächer machen mir Spaß.“)
- Angst (z. B. „Ich habe Angst vor den meisten Schulfächern.“)

Höhere Mittelwerte stehen hier für ein positiveres schulisches Selbstkonzept, ein größeres schulisches Interesse sowie für mehr schulbezogene Angst.

4.3 Klassen- und Schulklima

Wie Schülerinnen und Schüler ihre Schule beziehungsweise ihre Klasse wahrnehmen, wird anhand von drei Skalen erfasst:

- Lehrer-Schüler-Beziehung (z. B. „In meiner Klasse bekomme ich zusätzliche Hilfe von meinen Lehrern, wenn ich sie brauche.“)
- Klassenklima (z. B. „In meiner Klasse verstehen wir uns gut.“)
- Störneigung (z. B. „In meiner Klasse stören wir manchmal absichtlich den Unterricht.“)

Höhere Mittelwerte stehen hier für eine positivere Lehrer-Schüler-Beziehung, ein besseres Klassenklima, oder aber für eine ausgeprägtere Störneigung.

5 Referenzen

- MENFP (2007). *Français. Socle de compétences. Niveau fin 6^e ES*. Luxembourg: MENFP.
- MENFP (2008). *L'approche par compétences. Enseignement primaire. Cycle 2*. Luxembourg: MENFP.
- MENFP (2011). *Plan d'études. École fondamentale*. Luxembourg: MENFP.
- MENFP (2008a). *Bildungsstandards Sprachen*. Luxembourg: MENFP.
- MENFP (2008b). *Mathématiques. Division inférieure de l'enseignement secondaire. Compétences disciplinaires*. Luxembourg: MENFP.
- MENFP (2008c). *Mathématiques. Division inférieure de l'enseignement secondaire technique. Compétences disciplinaires*. Luxembourg: MENFP.