



ÉpStan – Fondamental (Zyklus 2.1)

Erläuterungen zum Klassen- und Schulbericht

1	Erklärung Kompetenzen	2
1.1	Was messen die ÉpStan? Was messen sie nicht?	2
1.2	Wie werden die <i>Socles</i> erfasst?	3
1.3	Was unterscheidet die ÉpStan von klassischen Prüfungen?.....	3
2	Wie zuverlässig messen die ÉpStan?	4
3	Auswertung der Testergebnisse	5
3.1	ÉpStan-Metrik	5
3.2	Was ist der Erwartungsbereich?	5
4	Erklärung Lernmotivation	6
4.1	Lernmotivation im Fach.....	7
4.2	Generelle schulische Lernmotivation.....	7
4.3	Klassen- und Schulklima.....	8
5	Referenzen	8

1 Erklärung Kompetenzen

1.1 Was messen die ÉpStan? Was messen sie nicht?

Die Kompetenzen der Schülerinnen und Schüler werden im Rahmen der ÉpStan in Form von standardisierten Tests erfasst. Die ÉpStan-Aufgaben, die im Rahmen der Kompetenztests eingesetzt werden, basieren auf den Bildungsstandards, welche vom MENJE für die verschiedenen Zyklen der Grund- und Sekundarschule entwickelt wurden. Kompetenzen sind in den Bildungsstandards wie folgt 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).

Derzeit bearbeiten die Schülerinnen und Schüler im Rahmen der ÉpStan im Zyklus 2.1 Aufgaben zur Erfassung der mathematischen Kompetenz, dem Hörverstehen in Luxemburgisch und den Vorläuferkompetenzen des Schriftspracherwerbs. Die Aufgaben variieren hinsichtlich ihrer Schwierigkeit. So kann etwa eine Mathematikaufgabe erfordern, dass von zwei Mengen von Äpfeln die größte Menge zu erkennen und anzukreuzen ist (geringe Schwierigkeit), oder aber, dass eine anhand von symbolischen Zahlen angegebene Addition zu lösen ist (höhere Schwierigkeit).

An dieser Stelle ist es nochmals wichtig zu betonen, dass die ÉpStan-Kompetenztests nicht alle in den Bildungsstandards definierten Kompetenzen überprüfen. Zum einen, da mit standardisierten Aufgabenformaten nicht alle Kompetenzbereiche erfasst werden können. Zum anderen aufgrund der begrenzten Zeit, die zur Durchführung der ÉpStan-Kompetenztests zur Verfügung steht. Aus diesen Gründen wird eine repräsentative Auswahl von möglichst ökonomisch und standardisiert erfassbaren Schlüsselkompetenzen gewählt:

- **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 Wortwiedererkennung.
- **Hörverstehen Luxemburgisch 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, Reportage, Hörspiel). Es werden Aufgaben zu verschiedenen Kompetenzen entwickelt (vgl. MENFP, 2011, S. 6), z. B. seinen Gesprächspartner verstehen, Informationen lokalisieren und verstehen, sowie Hörstrategien kennen und nutzen.

Wir sind uns bewusst, dass die hier gemessenen Sprachkompetenzen, nur zwei, wenn auch sehr wichtige Facetten sprachlicher Fertigkeiten darstellen. Während die mündliche Sprachproduktion auch eine wichtige sprachliche Kompetenz darstellt, wird

diese nicht erhoben da sie im Kontext der ÉpStan im Gegensatz zu den ausgewählten Kompetenzen nicht standardisiert, ökonomisch und objektiv erfasst werden kann.

Die ÉpStan liefern somit eine hochstandardisierte, aber unvollständige Momentaufnahme, während die *Bilans* auf einer langfristigen und vollständigen Evaluation beruhen. Das Ergebnis der ÉpStan versteht sich also komplementär und nicht in Konkurrenz zu den *Bilans*.

1.2 Wie werden die *Socles* erfasst?

In einer ersten Phase wird jede Aufgabe, unter strikter Berücksichtigung der theoretischen Deskriptoren (MENFP, 2011, S. 6, 8, 28-30), durch die Testentwicklungsgruppen einer theoretischen Schwierigkeitsstufe („unter *Niveau Socle*“, „*Niveau Socle*“, „*Niveau Avancé*“, „über *Niveau Avancé*“) zugeordnet. Diese Klassifikation erfolgt vor der Datenerhebung, d. h., bevor empirische Schwierigkeitsindikatoren für die Aufgaben vorliegen. Nach der Datenerhebung werden in einer zweiten Phase die theoretischen und empirischen Aufgabenschwierigkeiten miteinander verglichen. Wenn diese für eine Aufgabe nicht übereinstimmen, wird entweder eine theoretische Reklassifikation vorgenommen – sofern diese anhand der Lernstufendeskriptoren vertretbar ist – oder die besagte Aufgabe wird nicht in die Berechnung der *Socles* mit aufgenommen.

Durch die hier stark vereinfacht dargestellte Prozedur können die Testentwicklungsgruppen sicherstellen, dass theoretisch schwierige Aufgaben auch tatsächlich empirisch schwierig ausfallen und umgekehrt. Außerdem kann auf diese Art und Weise sichergestellt werden, dass sich alle Aufgaben, welche z. B. zur Berechnung des *Niveau Socle* in Mathematik benutzt werden, in etwa im gleichen empirischen Schwierigkeitsbereich bewegen und die Schwierigkeit des *Niveau Socle* somit relativ homogen ausfällt.

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 geht man davon aus, dass Schülerinnen und Schüler, die am Unterricht teilgenommen haben, alle Aufgaben (prinzipiell) lösen könnten. Sind alle Aufgaben richtig gelöst, erfüllen die Schülerinnen und Schüler die Grunderwartung, den Sollzustand. Wirklich beweisen können sich Schülerinnen und Schüler bei klassischen Prüfungen jedoch nicht, da die Aufgaben in der Regel nicht über den Sollzustand hinausgehen. Andererseits ist es für Schülerinnen und Schüler ohne Weiteres möglich, dass ihre Leistung unter dem Soll liegt. Diese Sichtweise ist sehr stark in unserer Prüfungskultur verankert. Dies zeigt sich vor allem auch daran, dass bei der Verbesserung von Prüfungen kaum darauf aufmerksam gemacht wird, was die Schülerinnen und Schüler richtig machen. Im Gegenteil: Die Korrektur hebt vor allem hervor, was Schülerinnen und Schüler (noch) falsch machen. Zusammenfassend ist demnach festzuhalten, dass es bei klassischen Prüfungen eher weniger darauf ankommt, was Schülerinnen und Schüler können, sondern vielmehr, ihre Schwächen bzw. ihre Abweichung vom Soll zu identifizieren.

Im Gegensatz hierzu sind die ÉpStan so ausgelegt, dass sie die komplette Kompetenzbandbreite abdecken, die man bei Schülerinnen und Schülern beobachtet:

Das was jeder können sollte (*Niveau Socle*) und das, was (weit) darüber hinausgeht (*Niveau Avancé* und höher). Dies hat zur Folge, dass man von vornherein nicht davon ausgeht, dass alle Schülerinnen und Schüler alle Aufgaben lösen können. Schülerinnen und Schüler können ihre Kompetenzen unter Beweis stellen, indem sie versuchen, möglichst viele Aufgaben zu lösen. Es kommt hier also nicht vorrangig darauf an, Schwächen zu identifizieren. Im Vordergrund steht festzustellen, was Schülerinnen und Schüler können.

Die Tatsache, dass das Können der Schülerinnen und Schüler im Vordergrund steht, zeigt, dass die ÉpStan explizit auf dem Prinzip des kompetenzorientierten Unterrichts beruhen. Letzteres heißt jedoch nicht automatisch, dass die Aufgaben der ÉpStan auch pauschal repräsentativ für kompetenzorientierten Unterricht sind. Der Sachverhalt ist etwas komplizierter.

Die ÉpStan zielen explizit auf die Vergleichbarkeit der Schülerleistungen untereinander sowie mit einem Standard (dem *Niveau Socle*) ab. Die getesteten Kompetenzbereiche wurden ausgewählt, weil sie von zentraler Bedeutung sind und sich gleichzeitig mit verhältnismäßig wenig Aufwand standardisiert, objektiv und zuverlässig erfassen lassen. Das Gleiche gilt für die ausgewählten Aufgabenformate. Unter Anbetracht dieser Zielsetzung, in Verbindung mit testökonomischen Gründen, stellen geschlossene und halboffene Aufgabenformate für die ÉpStan die beste Wahl dar. Der tagtägliche Unterricht setzt jedoch andere Prioritäten, da der Lernprozess der Schülerinnen und Schüler im Mittelpunkt steht. Dieser ist vorzugsweise mit offenen Aufgabenformaten zu fördern.

2 Wie zuverlässig messen die ÉpStan?

Zur Erfassung der Kompetenzen in der Grundschule wurden erstmals bei den ÉpStan 2014/15 im Zyklus 2.1 vier Testhefte (2-mal Sprachkompetenz, 2-mal Mathematikkompetenz) eingesetzt. Die Antworten der Schülerinnen und Schüler in den Tests wurden von ihren Lehrkräften korrigiert und in Antwortbögen übertragen. Hierfür erhielten die Lehrkräfte detaillierte Manuale, wie sie die Antworten der Schüler bei den Aufgaben in den Kompetenztests korrigieren sollen. Bei den ÉpStan ist ein zentrales Ziel, dass die Schülerkompetenzen so standardisiert wie möglich, d.h. objektiv und zuverlässig – unter Berücksichtigung der gegebenen Rahmenbedingungen – erfasst werden. Maßnahmen, wie die ausschließliche Anwendung von geschlossenen und halboffenen Aufgabenformaten, die Bereitstellung detaillierter Korrekturanweisungen (zur Vermeidung einer zu milden bzw. zu strengen Korrektur) sowie die computerbasierte Erfassung der Antwortbögen, 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 2015/16 eine Zusatzstudie im Zyklus 2.1 durchgeführt. Von jedem der vier Testhefte wurde 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 120 Schülerhefte von einem zweiten, hierfür geschulten Korrektor streng nach den im Manual vorgegebenen Korrekturanweisungen erneut verbessert. Insgesamt lagen so für 3000 Schülerantworten jeweils Verbesserungen von

zwei Korrektoren vor. Das wichtigste Ergebnis war, dass die Korrekturen in höchstem Maße übereinstimmten: Absolut lag die Übereinstimmung bei 97,2%; berücksichtigt man, dass die Übereinstimmung zwischen den Korrektoren auch zufallsbedingt sein könnte, lag die korrigierte Übereinstimmung immer noch bei 96,1%. 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.

3 Auswertung der Testergebnisse

3.1 ÉpStan-Metrik

Ein wichtiges Ziel der ÉpStan ist es, die Kompetenzen verschiedener Schülerjahrgänge über die Zeit hinweg zu vergleichen, um mögliche Effekte von Reformen im Bildungswesen sichtbar zu machen. Es können nicht in jedem Schuljahr die gleichen Tests eingesetzt werden, da ansonsten die Möglichkeit besteht, dass Lehrkräfte ihre Schülerinnen und Schüler gezielt auf den Test vorbereiten, was zu verzerrten Einschätzungen der Wirksamkeit von Reformprogrammen führen würde. Aus diesem Grund wurden die Testleistungen der Schülerinnen und Schüler nicht einfach zu Punktsummen zusammengefasst, sondern es wurden mit statistischen Modellen aus der sogenannten *Item Response Theory* (IRT) Kompetenzwerte berechnet. Diese Kompetenzwerte können über verschiedene Schülerjahrgänge hinweg verglichen werden, obwohl unterschiedliche Kompetenztests eingesetzt werden (hierfür muss nur ein Teil der Aufgaben wiederholt eingesetzt werden).

Die Kompetenzwerte werden so normiert, dass der Mittelwert über die gesamte Schülerschaft in Luxemburg für das Schuljahr 2014/15 bei 500 Punkten liegt. Die Standardabweichung (also die mittlere Abweichung der Testwerte vom Mittelwert) wurde auf 100 Punkte gesetzt. Das Schuljahr 2014/15 dient demnach als Referenzjahr. Das heißt, die diesjährigen ÉpStan-Ergebnisse sind vergleichbar bis hin zum Schuljahr 2014/15. Verändert sich also der mittlere ÉpStan-Punktwert einer Gruppe von Schülerinnen und Schülern von einem Jahr zum anderen, so deutet dies auf eine tatsächliche Veränderung der Kompetenzen in dieser Gruppe hin.

3.2 Was ist der Erwartungsbereich?

Viele Lehrkräfte wissen von vornherein, dass ihre Klasse aufgrund der soziokulturellen und sozioökonomischen Zusammensetzung ziemlich sicher unter oder über dem Landesdurchschnitt abschneiden wird. Warum soll man also überhaupt testen, wenn das Ergebnis sowieso von vornherein festzustehen scheint?

Zahlreiche internationale und nationale Studien (z. B. PISA, PIRLS, die ÉpStan) haben immer wieder gezeigt, dass die genannten Merkmale (Geschlecht, sozialer Hintergrund, Muttersprache) deutlich mit der Leistung in Kompetenztests zusammenhängen. So geht beispielsweise in Luxemburg mit weiblichem Geschlecht, luxemburgischer oder deutscher Muttersprache und hohem sozialen Status im Mittel eine höhere Leistung in einem Deutschttest einher. Andererseits haben Jungen mit portugiesischer oder französischer Muttersprache und niedrigem sozialen Status im Mittel eine eher schlechtere Leistung in einem Deutschttest.

Die soziokulturelle und sozioökonomische Zusammensetzung der Schülerschaft unterscheidet sich erheblich zwischen Klassen. Leistungen der Klassen sind also nicht ohne Weiteres zu vergleichen.

Am deutlichsten lässt sich dies wohl an einem (stark vereinfachten) Beispiel erläutern: Nehmen wir eine fiktive Schulklasse an, die einen Deutschttest absolviert hat. Die Schülerinnen und Schüler haben dabei ein Ergebnis von im Mittel 430 Punkten erreicht (die Punktabgaben beziehen sich auf die ÉpStan-Metrik aus Abschnitt 3.1) und liegen somit deutlich unter dem nationalen Durchschnitt. Nehmen wir weiter an, dass die Klasse vor allem von Jungen aus sozial weniger gut gestelltem Elternhaus besucht wird, die zu Hause Portugiesisch oder Französisch sprechen. Vor diesem Hintergrund ist das unterdurchschnittliche Abschneiden nicht überraschend. Die interessantere Frage lautet aber: wie verhält sich das tatsächliche Ergebnis der Klasse zu demjenigen, das aufgrund der Zusammensetzung der Schülerschaft ohnehin zu erwarten gewesen wäre? Um diese Frage zu beantworten, berechnen wir „Erwartungsbereiche“ anhand eines statistischen Modells.

Dabei gehen wir wie folgt vor: Mithilfe eines statistischen sogenannten Regressionsmodells ermitteln wir zunächst den Zusammenhang von Testleistungen in den gemessenen Kompetenzbereichen einerseits und Geschlecht, sozialem Hintergrund, sprachlichem und Migrationshintergrund, sowie Schullaufbahndaten andererseits. Die genannten Merkmale bezeichnen wir als Hintergrundvariablen.

Im zweiten Schritt verwenden wir die Parameter des Regressionsmodells, um für jede Klasse einen Erwartungsbereich zu berechnen, in dem das Ergebnis der Klasse liegen sollte, wenn die im Regressionsmodell abgebildeten Zusammenhänge dort ebenso gelten wie im gesamten Land. Die Breite dieses Erwartungsbereichs spiegelt die Messgenauigkeit der eingesetzten Tests, die Vorhersagegenauigkeit des Regressionsmodells und 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 sozioökonomischer Merkmale der Schülerinnen und Schüler statistisch zu erwarten gewesen wäre. Ein Ergebnis unterhalb des Erwartungsbereichs signalisiert jedoch ein schlechteres Resultat, als statistisch zu erwarten gewesen wäre.

Indem das Leistungsprofil der eigenen Klasse also in Bezug zu Klassen mit vergleichbarer soziokultureller und sozioökonomischer Zusammensetzung gesetzt wird, können auch Lehrkräfte, welche z. B. sehr effektiv unter „schwierigen“ Bedingungen arbeiten, eine positive und wissenschaftlich abgesicherte Leistungsrückmeldung zu ihrer Klasse erhalten. Erwartungsbereiche sind jedoch nicht nur für Lehrkräfte, welche unter „schwierigen“ Bedingungen arbeiten, interessant, sondern für alle Lehrkräfte, die sich ein faires Bild der Schülerleistungen ihrer Klassen machen wollen.

4 Erklärung Lernmotivation

Zusätzlich zu den Leistungstests haben die Schülerinnen und Schüler einen Fragebogen bearbeitet. Dieser erfragte folgende Bereiche:

- Lernmotivation in den ÉpStan geprüften Fächern

- generelle Lernmotivation in der Schule
- Klassen- und Schulklima

Die Lernmotivation der Schülerinnen und Schüler wurde im Fragebogen mittels Aussagen wie „Ich interessiere mich für die meisten Schulfächer“ erfasst. Durch Ankreuzen gaben die Schülerinnen und Schüler den Grad ihrer Zustimmung zu diesen Aussagen an. Letzteres erfolgte auf einer zweistufigen Antwortskala mit den Antwortkategorien „trifft nicht zu“ oder „trifft zu“. Die Antworten der verschiedenen Aussagen haben wir dann thematisch zu Skalenwerten zusammengefasst (z. B. eine Skala für generelles Schulinteresse).

4.1 Lernmotivation im Fach

Die Lernmotivation der Schülerinnen und Schüler im Fach wurde anhand von drei Skalen erfasst: Selbstkonzept, Interesse und Angst.

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

Höhere Mittelwerte bedeuten positiveres Selbstkonzept, größeres Interesse und mehr Angst in diesem Fach.

4.2 Generelle schulische Lernmotivation

Die generelle schulische Lernmotivation der Schülerinnen und Schüler wurde 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 bedeuten positiveres Selbstkonzept, größeres Interesse und mehr Angst.

4.3 Klassen- und Schulklima

Wie Schülerinnen und Schüler ihre Schule beziehungsweise ihre Klasse wahrnehmen, wurde 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 bedeuten positivere Lehrer-Schüler-Beziehung, positiveres Klassenklima und höhere 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.