



LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Ministère de l'Éducation nationale,
de l'Enfance et de la Jeunesse



UNIVERSITY OF LUXEMBOURG
Luxembourg Centre for
Educational Testing (LUCET)



ÉpStan – Secondaire (5^e)

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 Lernstufen erfasst?	4
1.3	Was unterscheidet die ÉpStan von klassischen Prüfungen?	4
2	Auswertung der Testergebnisse.....	5
2.1	ÉpStan-Metrik	5
2.2	Was ist der Erwartungsbereich?	6
3	Erklärung Lernmotivation	7
3.1	Lernmotivation im Fach	7
3.2	Generelle schulische Lernmotivation.....	8
3.3	Klassen- und Schulklima.....	8
4	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).

Folgt man dieser Definition, werden Kompetenzen evaluiert, indem Schülerinnen und Schüler Aufgaben mit komplexen Problemstellungen bearbeiten. Für die adäquate Lösung eines solch komplexen Problems müssen die Schülerinnen und Schüler ihr Wissen und ihr Können integrativ einsetzen. Die Bearbeitung eines Problems ist meist zeitintensiv und anstrengend und erfordert deshalb von den Schülerinnen und Schülern auch ein hohes Maß an Motivation.

Was bedeutet „komplex“ im Zusammenhang mit Problemen? Im Rahmen der ÉpStan unterscheiden wir verschiedene Grade an Komplexität. Die Komplexität kann dabei innerhalb eines Faches variieren (z. B. Grundrechenaufgaben vs. Aufgaben zum mathematischen Problemlösen), aber auch dadurch, dass interdisziplinäre (fächerübergreifende) Probleme gestellt werden (in denen z. B. eine Schreibaufgabe auf einer naturwissenschaftlichen Problemstellung aufbaut). Derzeit bearbeiten die Schülerinnen und Schüler im Rahmen der ÉpStan in der 5^e nur fachspezifische Probleme zur Erfassung der mathematischen Kompetenz und der Lesekompetenz in Deutsch und Französisch. Die Probleme variieren hinsichtlich der Komplexität ihrer Anforderungen. So kann etwa eine Leseaufgabe erfordern, dass eine explizit im Text erwähnte Information zu lokalisieren ist (geringe Komplexität), oder aber, dass eine komplexe Schlussfolgerung aufgrund textinhärenter Informationen zu ziehen ist (höhere Komplexität). Interdisziplinäre Probleme sind derzeit nicht in den Kompetenztests der ÉpStan enthalten.

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 in der 5^e.** Die Mathematikaufgaben in der 5^e fokussieren auf mathematisches Modellieren, bzw. Problemlösen, wobei sich die Aufgaben und Probleme deutlich in ihrem Komplexitätsgrad unterscheiden. Die Aufgaben stammen dabei aus einem von drei mathematischen Inhaltsbereichen: (1) Zahlen und Operationen (MENFP 2008b; 2008c, S. 21-23), (2) Ebene und räumliche Figuren (MENFP 2008b; 2008c, S. 17-19) und (3) Abhängigkeit und Veränderung

(MENFP 2008b; 2008c, S. 25-27). Die Schülerinnen und Schüler können die Mathematikaufgaben jeweils auf Deutsch oder Französisch bearbeiten.

- **Leseverstehen Deutsch 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 Unterkompetenzen sind nicht explizit in den Bildungsstandards vorgesehen, lassen sich jedoch aus der Beschreibung der vier Lernstufen ableiten (vgl. MENFP 2008a, S. 68).

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

- **Leseverstehen Französisch in der 5^e.** Der Test zur Erfassung der Lesekompetenz in Französisch basiert wie der Deutsch-Lesetest 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). Diese Unterkompetenzen sind nicht explizit in den Bildungsstandards vorgesehen, lassen sich jedoch aus der Beschreibung der vier Lernstufen ableiten (vgl. MENFP 2008a, S. 68). Jeder 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 – voies d’adaptation et de préparation“ werden unterschiedliche Testversionen erstellt, die stets eine Mindestanzahl an sich überschneidendem Text- und Aufgabenmaterial enthalten, um ihre Vergleichbarkeit zu gewährleisten.

Je nach der Schulform, die eine Schülerin oder ein Schüler besuchte, bearbeitete sie oder er eine im Schwierigkeitsgrad angepasste Version der Tests. Unterschieden wurde zwischen drei Versionen:

- *Enseignement secondaire classique (ESC)*
- *Enseignement secondaire général – voie d’orientation (ESG)*
- *Enseignement secondaire général – voies d’adaptation et de préparation (ESG-P)*

Die Testdauer betrug für alle Schülerinnen und Schüler, unabhängig von der Testversion, 45 Minuten pro Fach, wobei die Reihenfolge der Tests variierte. Eine Mindestanzahl an Text- und Aufgabengabenmaterial ist schulformübergreifend in allen Testversionen enthalten. Jede Testversion enthält aber auch viele Aufgaben, die auf den Leistungsbereich der Schülerschaft der jeweiligen Schulform abgestimmt sind. Durch die Verwendung gemeinsamer Aufgaben über Schulzweige hinweg sowie das Zurückgreifen auf moderne psychometrische Verfahren, ist ein direkter Kompetenzvergleich über die Schulzweige hinweg gesichert.

1.2 Wie werden die Lernstufen erfasst?

In einer ersten Phase wird jede Aufgabe, unter strikter Berücksichtigung der theoretischen Deskriptoren (MENFP 2008a, S. 68; MENFP 2008b; 2008c), durch die Testentwicklungsgruppen einer theoretischen Schwierigkeitsstufe 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 Lernstufen 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 der Lernstufe 2 in Leseverstehen Französisch benutzt werden, in etwa im gleichen empirischen Schwierigkeitsbereich bewegen und die Schwierigkeit der Lernstufe 2 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 darauf, 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 fast jeder kann, das was jeder (mindestens) können sollte und das, was (weit) darüber hinausgeht. 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 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 Auswertung der Testergebnisse

2.1 ÉpStan-Metrik

Ein wichtiges Ziel der ÉpStan ist es, die Kompetenzen verschiedener Schülerjahrgänge und Schulformen 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 Punktsammenfassungen zusammengefasst, sondern es wurden mit statistischen Modellen aus der sogenannten *Item Response Theory* (IRT) Kompetenzwerte berechnet. Diese Kompetenzwerte können über verschiedene Schülerjahrgänge und Schulformen hinweg verglichen werden, obwohl unterschiedliche Kompetenztests eingesetzt werden (hierfür muss nur ein Teil der Aufgaben wiederholt eingesetzt werden).

Die Kompetenzwerte wurden so normiert, dass der Mittelwert über die gesamte Schülerschaft (d.h. über alle Schulformen) in Luxemburg für das Schuljahr 2011/12 bei 500 Punkten liegt. Die Standardabweichung (also die mittlere Abweichung der Testwerte vom Mittelwert) wurde auf 100 Punkte gesetzt. Das Schuljahr 2011/12 dient demnach als Referenzjahr. Das heißt, die diesjährigen ÉpStan-Ergebnisse sind vergleichbar bis hin zum Schuljahr 2011/12. 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.

2.2 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 Deutschtest einher. Andererseits haben Jungen mit portugiesischer oder französischer Muttersprache und niedrigem sozialen Status im Mittel eine eher schlechtere Leistung in einem Deutschtest.

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 Deutschtest absolviert hat. Die Schülerinnen und Schüler haben dabei ein Ergebnis von im Mittel 430 Punkten erreicht (die Punktangaben beziehen sich auf die ÉpStan-Metrik aus Abschnitt 2.1) 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 Mathematik und Deutsch (sowie Französisch in der Sekundarstufe) einerseits und Geschlecht, sozialem Hintergrund, sprachlichem und Migrationshintergrund, Schullaufbahndaten, sowie Schulform (in der Sekundarstufe) andererseits. Die genannten Merkmale bezeichnen wir als Hintergrundvariablen.

Im zweiten Schritt verwenden wir die Parameter des Regressionsmodells, um für jede Klasse einen Wertebereich 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 interessant, welche unter „schwierigen“ Bedingungen arbeiten, sondern für alle Lehrkräfte, die sich ein faires Bild der Schülerleistungen ihrer Klassen machen wollen.

3 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 vierstufigen Antwortskala mit den Antwortkategorien „trifft nicht zu“ bis „trifft zu“. Die Antworten der verschiedenen Aussagen haben wir dann thematisch zu Skalenwerten zusammengefasst (z. B. eine Skala für generelles Schulinteresse). Die Schülerinnen und Schüler konnten den Fragebogen wahlweise auf Deutsch oder Französisch bearbeiten.

3.1 Lernmotivation im Fach

Die Lernmotivation der Schülerinnen und Schüler im Fach wurde anhand von vier Skalen erfasst: Selbstkonzept, Interesse, Angst und Leseinteresse (nur in den Sprachfächern).

- 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 Deutsch.“
- 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 Mathematik.“
- 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 Deutsch.“

- Die Skala Leseinteresse gibt an, inwiefern die Schülerinnen und Schüler gerne lesen. Eine Beispielaussage hierfür ist: „Ich lese gerne Bücher und Zeitschriften in französischer Sprache.“

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

3.2 Generelle schulische Lernmotivation

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

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

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

3.3 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 untereinander 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.

4 Referenzen

- MENFP (2007). *Français. Socle de compétences. Niveau fin 6^e ES*. 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.