

Evaluation commune : glossaire alphabétique

Auteur : Romano Mero, chef de projet « évaluation commune » au WBZ CPS, juillet 2011

Conseil : Prof. Franz Baeriswyl, Université de Fribourg

Cadre européen commun de référence pour les langues (CECR)	Cf. <i>cadre de référence</i> .
Cadre de référence	Décrit précisément quelles capacités les élèves devraient avoir dans un domaine et quelles situations-problèmes devraient être résolues (formulations « sait faire »). Base pour la planification et le développement de programmes d'apprentissage des langues et de diplômes de langue. Les <i>compétences</i> et le niveau requis sont définis à l'aide de descripteurs empiriquement établis. Exemple : <i>Cadre européen commun de référence pour les langues</i> (CECR).
Caractéristiques de la procédure	Caractéristiques de la procédure lors de l'élaboration, du passage et de l'évaluation d' <i>épreuves communes</i> (p. ex. degré d'obligation, moment du passage de l'épreuve, champ de l'épreuve, etc.).
Certification	Une épreuve standardisée au niveau national ou international et dont les résultats sont 100% comparables (exemples: le diplôme DELF pour le français, le certificat Cambridge pour l'anglais et l'examen PLIDA pour l'italien).
Compétences	Capacités et aptitudes cognitives innées ou acquises permettant à l'individu de résoudre certains problèmes ; capacités et aptitudes motivationnelles, volitives et sociales qui en découlent et permettant d'utiliser avec succès et de manière responsable ces solutions dans différentes situations (d'après Weinert). Avoir des compétences dans un domaine particulier signifie que l'on peut agir avec succès dans ce domaine (d'après Klieme, Maag-Merki, Hartig).
Compétence diagnostique	Capacité qu'a l'enseignant de préparer et d'effectuer une évaluation des performances permettant de tirer les bonnes conclusions en ce qui concerne les processus d'enseignement et d'apprentissage.
Compétences fondamentales (en vue d'études supérieures)	Compétences indispensables (RRM) en ce qui concerne l'aptitude générale à étudier.
Conditions cadre	Directives pour l'élaboration et la réalisation de l'épreuve, p. ex. définition des <i>domaines de contenus</i> et de <i>compétences</i> , des outils de travail autorisés, etc.
Confidentialité	Une des <i>exigences liées à la procédure</i> : le niveau de confidentialité peut être faible ou élevé suivant la manière dont l'épreuve est corrigée (p. ex de manière anonyme) et dont les résultats sont traités (communication aux élèves et à d'autres personnes).
Critères d'évaluation	Les critères qui permettront d'évaluer une épreuve peuvent être fixés à l'avance, en se référant par exemple à des réponses types.
Critères de qualité	Egalement appelés critères de qualité des tests : <i>objectivité, validité, fiabilité</i> sont les 3 principaux critères. Critères supplémentaires : <i>égalité des chances, rendement, fair-play</i> .
Développement individuel	<i>L'évaluation formative des performances</i> vise le développement individuel de l'élève.

Directives de correction	Directives élaborées ou prescrites en commun et indiquant comment une épreuve doit être corrigée et évaluée.
Domaine de contenus	Les domaines de contenus sont fixés d'après les directives des <i>plans d'études</i> cantonaux en fonction des performances à évaluer. On a d'une part les domaines de contenus (définis par un niveau élevé d' <i>harmonisation</i> et une liberté restreinte de l'enseignant lors de l'établissement de son programme) et d'autre part les <i>domaines de compétences</i> (cf. <i>modèle argovien</i>).
Domaine de compétences	Formulation des exigences que les élèves doivent satisfaire dans un domaine d'une discipline (cf. modèle d'harmonisation 4 de Baeriswyl). On a d'une part la détermination des domaines de compétences dans une discipline et d'autre part celle des domaines de contenus.
Domaine d'exigences	Précise le type et la <i>pondération</i> de la performance attendue à l'épreuve. Exemples de domaines d'exigences (d'après Bloom) : <ul style="list-style-type: none"> - Connaître, savoir et comprendre : exécuter des procédures courantes suivant des directives, appliquer des formules et règles élémentaires (niveau I). - Appliquer, utiliser : choisir et appliquer des stratégies simples pour résoudre les problèmes, travailler avec des modèles (niveau II). - Jugement scientifique, créativité : concevoir des modèles pour des situations complexes et les utiliser dans de nouvelles situations (niveau III). Chaque domaine d'exigences peut compter plusieurs <i>niveaux d'exigences</i> .
Egalité des chances	Une épreuve doit mesurer une performance indépendamment de l'école, de la classe et du contexte socio-économique ; c'est un des <i>critères de qualité d'un test</i> .
Epreuve identique	Le <i>sujet</i> , les exercices, la <i>pondération</i> , les <i>conditions cadre</i> et les <i>exigences liées à la procédure</i> sont les mêmes. Une épreuve identique correspond au plus haut degré d' <i>harmonisation</i> .
Epreuve commune	Epreuve en tant que telle, résultat du processus d' <i>évaluation commune</i> (mené par un groupe de discipline).
Evaluation commune	Processus de discussion (au sein d'un groupe de discipline) pour élaborer une <i>épreuve commune</i> . Evaluer en commun signifie formuler des attentes communes relatives à un <i>domaine d'exigences</i> précis.
Epreuve comparable	Epreuve qui, grâce à son niveau d' <i>harmonisation</i> par rapport à une autre épreuve, permet une comparaison des performances obtenues.
Epreuve « de promotion »	L'épreuve permet-elle la promotion au degré supérieur ou non ? (cf. <i>évaluation sommative des performances</i>)
Epreuve transversale	Synonyme d' <i>épreuve commune</i> . Peut p. ex. comprendre toute une volée d'élèves (dans une ou plusieurs école-s).
Evaluation formative des performances	Evaluation visant l'amélioration et le développement individuel des élèves pendant leur formation. Analogie au club de sport : l'entraîneur conseille ses protégés pour que chacun s'entraîne de manière optimale et puisse s'améliorer.
Evaluation pronostique des performances	Evaluation concernant une performance future. Analogie au club de sport : l'entraîneur choisit ses meilleurs sportifs pour le prochain championnat. Voir aussi <i>évaluation sommative</i> et <i>évaluation formative</i> des performances.
Evaluation sommative des performances	Evaluation finale visant à contrôler le <i>taux de réussite</i> d'un objectif d'apprentissage. Analogie au club de sport : l'entraîneur veut savoir à quel niveau de performances ses joueurs se situent et fixe une date (critère

	temporel) à laquelle un objectif doit être atteint.
Exigences liées à la procédure	Exigences concernant l'élaboration, le passage et l'évaluation d'épreuves communes. Elles sont développées sur la base des <i>caractéristiques de la procédure</i> d'évaluation commune.
Fair-play	L'épreuve doit être claire ; en d'autres termes, ses objectifs, sa forme et son <i>sujet</i> doivent être communiqués aux élèves de manière claire, univoque et égale pour tous ; c'est un des <i>critères de qualité</i> d'un test.
Fiabilité	Un des trois principaux <i>critères de qualité</i> des tests tirés de la <i>théorie des tests</i> : le test doit être cohérent. La fiabilité d'un test se mesure d'après les critères suivants : <ul style="list-style-type: none"> - Forme du test (la plus homogène possible) <li style="padding-left: 20px;"><i>Sélectivité</i> de chaque exercice - Indépendance des exercices les uns par rapport aux autres - Abandon des „speed-tests“ (connaissances liées à la vitesse) - Nombre d'items (exercices) - Nombre d'exercices de niveau moyen de difficulté .
Groupe de référence	Groupe dans lequel, resp. par rapport auquel une performance est mesurée (p. ex. une classe, une volée). Dans le cas d'une épreuve reposant sur la <i>norme sociale</i> , le <i>groupe de référence</i> est défini à l'avance.
Harmonisation/ Modèles d'harmonisation	Processus visant à rendre comparables les <i>épreuves communes</i> . Peut aller d'une <i>harmonisation superficielle</i> , de très faible niveau (outils de travail autorisés, durée, forme des questions) à une harmonisation à 100% (ex. épreuve identique). Voir les 6 modèles d'harmonisation selon Baeriswyl.
Harmonisation superficielle	Niveau d' <i>harmonisation</i> le plus faible. On définit p. ex. les outils de travail autorisés, la durée de l'épreuve et le type de questions.
Liberté pédagogique	Liberté qu'ont les enseignants de choisir leur programme et leur méthodologie (le processus d'évaluation commune n'entrave pas cette liberté).
Maturité interne	Epreuve de maturité interne à l'établissement et identique pour l'ensemble des élèves d'une discipline (cf. <i>épreuve identique</i>).
Modalités de passage de l'épreuve	Comment l'épreuve se déroule-t-elle (lieu, moment, outils autorisés) ? Ce type de dispositions correspond à une <i>harmonisation superficielle</i> et constitue un exemple d' <i>exigence liée à la procédure</i> .
Modèle argovien	Pour harmoniser les épreuves de maturité, le canton a fixé des <i>conditions cadre</i> unifiées aussi bien transversales que spécifiques, ainsi que des procédures et des directives d'examen uniformisées. Cela doit permettre d'augmenter la comparabilité des épreuves de maturité entre différentes écoles.
Modèle de compétences	Ensemble de compétences à tester dans une discipline (cf. p. ex. le <i>modèle argovien</i>).
Niveau d'exigences	<i>Taux de réussite</i> attendu par exercice. Le niveau d'exigences peut reposer sur un <i>modèle de compétences</i> ou un <i>plan d'études</i> . Il peut être normatif si la procédure prévoit une sélection. S'il est de caractère formatif, il permet de poser un diagnostic pour formuler d'autres exercices d'apprentissage. Il est important que le niveau d'exigences soit « correct » et judicieux pour l'ensemble des exercices et que l'on ne teste par exemple pas seulement les connaissances et les savoirs (cf. <i>domaine d'exigences</i>).

Niveau de performances	Qualité des performances par rapport à un groupe ou à un <i>niveau d'exigences</i> donné.
Norme basée sur une liste de critères	Le résultat indique si l'épreuve – sans tenir compte du groupe – est réussie ou non (<i>vs norme sociale</i>).
Norme sociale	Le résultat de l'épreuve indique le niveau de performance par rapport à un <i>groupe de référence</i> .
Objectivité	Un des trois principaux <i>critères de qualité</i> tirés de la <i>théorie des tests</i> . Les résultats des tests doivent être valables indépendamment de l'examineur et du contexte d'examen. En d'autres termes, lors du passage, de l'évaluation et de l'interprétation de l'épreuve, les examinateurs doivent éviter toute action subjective telle que reformulation, encouragement etc.
Plan d'études	Un des outils à disposition pour l' <i>harmonisation</i> des épreuves: p. ex. <i>plan d'études</i> cadre pour les écoles de maturité, principaux objectifs de l'école, objectifs par discipline, etc.
Pondération	Lors de l'évaluation, on définit l'importance de chaque exercice par rapport à l'ensemble de l'épreuve. On tient compte du programme/ <i>sujet</i> et du <i>niveau d'exigences</i> . C'est une des <i>exigences liées à la procédure</i> .
Rendement	La gestion du temps doit être optimale lors de l'élaboration d' <i>épreuves communes</i> ; il doit donc y avoir un juste équilibre entre temps investi et charge de travail ; c'est un des <i>critères de qualité</i> des tests.
Répartition des exercices	Une des <i>exigences liées à la procédure</i> , également nommée représentativité : il s'agit de définir comment les différents exercices sont répartis par <i>sujet</i> et s'ils sont représentatifs des objectifs d'apprentissage et du thème abordé. Cf. aussi <i>pondération</i> des exercices. .
Sélectivité	Lien entre l'exercice (item) et le résultat global de l'épreuve. Une sélectivité élevée signifie qu'un exercice permet, lorsqu'on regarde le résultat global, de différencier les élèves (les élèves connaissant bien les caractéristiques de l'épreuve résolvent « correctement » l'exercice, mais pas les autres élèves).
Sujet	<i>Domaines de contenus</i> testés dans une discipline.
Taux de réussite	Le taux de réussite concerne le <i>niveau d'exigences</i> d'une épreuve et s'exprime en pourcent. Ainsi, ce taux doit être de 50% pour qu'un examen de diplôme soit considéré comme réussi, alors qu'il doit souvent être de 66% à 75% pour obtenir la mention « suffisant » (note 4) à une épreuve gymnasiale.
Type d'exercice	On opère une distinction entre les exercices « ouverts » et « fermés ». Les exercices « ouverts » requièrent des critères d'évaluation et des directives de corrections détaillées. Les exercices « fermés », tels que les questionnaires à choix multiples, sont compliqués à créer mais plus simples à évaluer.
<i>Teaching to the test</i>	Préparation à l'épreuve et entraînement ciblé sur une épreuve et sur sa forme. Dans des cas extrêmes, cela peut donner lieu à du drill : exercices répétitifs visant à la restitution mécanique des connaissances.
Test	Ici synonyme d'épreuve.
Théorie des tests	Théorie qui permet de définir la qualité des épreuves (cf. <i>critères de qualité</i>).
Transparence	Une épreuve est transparente si elle est comparable à d'autres épreuves, p. ex. lorsque des <i>conditions cadre</i> , des <i>exigences liées à la procédure</i> , des <i>compétences</i> ou un programme ont été définis.

Validité	<p>Un des trois principaux <i>critères de qualité</i> des tests tirés de la <i>théorie des tests</i> : la mesure est valide lorsqu'elle mesure ce qui doit être mesuré :</p> <ul style="list-style-type: none">- Validité de contenu : l'objet de la mesure peut clairement être décrit.- Validité de critère : il y a un lien entre la performance et le critère avec lequel le test est corrélé.- Validité de construit : le test mesure l'aptitude qu'il doit évaluer et pas une autre.- Validité apparente : un profane voit le lien entre les exercices et l'objet évalué. <p>(D'après Urs Moser)</p>
----------	---

Bibliographie

- Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.
- Florek, H-Ch. (1999). Leistungsbegriff und Pädagogische Praxis. Münster: LIT
- Antonietti Ph. & Guignard, N. (2005) Mathematik. In: BFS/EDK. Bildungsmonitoring Schweiz. PISA 2003: Kompetenzen für die Zukunft. Zweiter nationaler Bericht. S. 17 - 34. Neuchâtel, Bern.
- Grunder, H-U./Bohl, Th. (Hrsg.) 2001. Neue Formen der Leistungsbeurteilung in den Sekundarstufen I und II. Hohengehren: Schneider.
- Hartig, J & Klieme, E. (2006) Kompetenz und Kompetenzdiagnostik. In Schweizer K. Leistung und Leistungsdiagnostik. Heidelberg: Springer.
- Jürgens, E. & Sacher, W. (2000). Leistungserziehung und Leistungsbeurteilung. Neuwied, Kriftel: Luchterhand.
- Klieme, E. Maag-Merki, K. & Hartig, J. (2007). Kompetenzbegriff und Bedeutung von Kompetenzen im Bildungswesen. In J. Hartig & E. Klieme (Hrsg.).
- Maier, U. (2010). Effekte testbasierter Rechenschaftslegung auf Schule und Unterricht. In: Zeitschrift für Pädagogik, 56, 1, S. 112 128.
- Moser, U. (2007). Gemeinsame Prüfungen auf testtheoretischer Basis entwickeln, Referat vom 6.11.2007 anlässlich des WBZ-Kurses 07.28.07.
- Neumann, M., Nagy, G., Trautwein, U., Lüdtke O. (2009) Vergleichbarkeit von Abiturleistungen. In: Zeitschrift für Erziehungswissenschaften, 12, S. 691–714.
- Rupp, A.A., Leucht, R., Hartung, R. (2006) „Die Kompetenzbrille“ aufsetzen – Verfahren zur multiplen Klassifikation von Lernenden für die Kompetenzdiagnostik in unterricht und Testung. In: Unterrichtswissenschaft, Zeitschrift für Lernforschung, 34. Jg. Heft 3, S. 195-219.
- Weinert, F. E. (2001). Leistungsmessungen in Schulen. Weinheim und Basel: Beltz
- Baumert et al. In Weinert, 2001. PISA-Zielsetzung.