

L'évaluation de la résolution de problèmes dans un cours de Mécanique pour ingénieurs : une comparaison entre deux méthodes

Maxim Morin¹, Thomas Gervais², David Ménard²,
david.menard@polymtl.ca

¹ Université de Montréal

² École Polytechnique de Montréal

RESUME : Les programmes de formation au génie accordent beaucoup d'attention au développement de la capacité des étudiants à résoudre des problèmes, mais rares sont les travaux de recherche qui se sont penchés sur l'évaluation de cette compétence. Il faut constater que l'évaluation dans les cours centrés sur la résolution de problèmes suit généralement une approche classique, où l'on demande aux correcteurs d'élaborer leur barème de correction. Pourtant, d'autres stratégies telles que le recours à des grilles à échelles descriptives sont propices à soutenir cette démarche de correction, et ce dans le but de rendre l'évaluation plus juste et équitable. Cette étude consiste d'une part à développer des grilles de correction pour un cours de Mécanique pour ingénieurs et d'autre part, à comparer la méthode de correction classique et la méthode nécessitant des grilles. Lors de la phase d'expérimentation, quatre correcteurs ont corrigé des échantillons identiques de solutions d'étudiants selon les deux méthodes, puis ils ont été interviewés. Les résultats de la correction montrent que l'utilisation des grilles ne réduit pas la variabilité du degré de sévérité des correcteurs, alors que les données des entretiens signalent que le choix d'une méthode de correction n'influence pas forcément les interprétations des correcteurs qui ne repèrent pas les mêmes erreurs et ne leur accordent pas la même importance. En définitive, les grilles à échelles descriptives forment un bon point de départ pour harmoniser les pratiques de correction, mais elles doivent être accompagnés d'une formation et d'un soutien appropriés.

Mots clés : résolution de problèmes, évaluation, grille à échelle descriptive, mécanique.

INTRODUCTION

Au quotidien, le travail de l'ingénieur porte sur l'analyse de situations problèmes variées et sur la proposition de solutions pour y remédier. Dans la circonstance, il est attendu que les futurs ingénieurs développent leur compétence à résoudre des problèmes variés au courant de leur formation professionnelle. Au Canada, le Bureau canadien d'agrément des programmes de génie (BCAPG) qui a le mandat d'orienter et de superviser les programmes de formation statue que les écoles et les facultés de génie ont la responsabilité de développer chez les étudiants douze qualités [1], desquelles fait partie la résolution de problèmes. Le BCAPG exige d'ailleurs aux programmes de démontrer qu'ils soient en mesure d'évaluer convenablement cette compétence.

Il faut pourtant remarquer que l'évaluation de la résolution de problèmes peut poser des défis considérables, notamment lorsque le cours est offert à de grands groupes, comme c'est souvent le cas dans la première année de formation. Ces futurs ingénieurs, qui suivent des cours d'un tronc commun au début de leurs études afin d'acquérir une base de connaissances et d'habiletés en génie, sont amenés dans le cadre de leurs activités d'apprentissages (les cours, les devoirs, les travaux pratiques) à résoudre des problèmes avec applications numériques dans divers domaines du génie et, par souci de cohérence entre l'enseignement et

l'évaluation, ils doivent passer des examens composés de tâches similaires.

Malgré les efforts pour réunir les conditions nécessaires à la passation d'examen de résolution de problèmes, le personnel enseignant se bute fréquemment à des défis de taille en ce qui concerne la correction des copies des étudiants. Dans les grandes classes, il faut inévitablement faire appel aux services de plusieurs correcteurs, ce qui mobilise des ressources organisationnelles et financières relativement importantes afin d'assurer une évaluation juste et équitable. Qui plus est, les enseignants n'ont pas toujours les outils nécessaires pour seconder le travail des correcteurs qu'ils embauchent.

Même si la recherche a accordé beaucoup d'intérêt à la résolution de problèmes, on recense peu de recherche qui a porté une attention particulière à l'évaluation de cette compétence et encore moins aux instruments d'évaluation de la résolution de problèmes [2]. Plus encore, il apparaît que les rares propositions semblent difficiles à utiliser ou non éprouvées empiriquement [3]. C'est faute de moyen que les enseignants s'en tiennent à une approche de correction « classique », pour laquelle ils demandent aux correcteurs d'élaborer individuellement leur barème de notation. Les correcteurs sont ainsi appelés à partager les points accordés entre les différents éléments des réponses manifestées dans les copies d'étudiants.

C'est en partie sur le constat d'une instrumentation insuffisante pour assurer une évaluation juste et équitable que cette étude a émergé. Il apparaît pourtant essentiel de vérifier la faisabilité de l'utilisation de nouveaux outils et leurs propriétés docimologiques. Le premier objectif de cette étude consiste à développer et à expérimenter un outil d'évaluation permettant d'harmoniser les pratiques de correction. Le second objectif est de comparer la méthode de correction dite « classique » à la méthode de correction recourant à ces outils..

LES GRILLES A ECHELLES DESCRIPTIVES

Jonassen [4], qui a étudié longuement la résolution de problèmes, propose de recourir à des grilles à échelles descriptives (scoring rubrics) pour l'évaluation de cette compétence. La grille à échelle descriptive est un outil de correction ou de notation des productions des étudiants. Elle sert généralement à apprécier la qualité de productions, de performances complexes ou des démarches, au regard de critères préalablement définis [5]. En comparaison à des démarches de correction « classique », il est soutenu que son utilisation permet en règle générale de réduire la subjectivité dans le processus de notation et de réduire la variabilité des interprétations entre les correcteurs [6, 7].

D'après Moskal [5], on distingue différents types de grilles à échelles descriptives. La grille du type global est formée d'une seule échelle de descripteurs et elle sert principalement à porter un jugement en considérant l'ensemble de la production de l'apprenant. Quant à elle, la grille du type analytique est constituée de plusieurs échelles descriptives associées à chacune des composantes de la performance. Elle cherche à décortiquer la production en dimensions indépendantes. Les grilles à échelles descriptives peuvent également se différencier entre la grille orientée vers une tâche spécifique et la grille dite générale qui peut s'adapter à un ensemble de performances similaires.

DEMARCHE

Le recours à des grilles à échelles descriptives a été le point de départ de ce projet de recherche qui s'est divisé en deux phases, le développement des grilles et l'expérimentation.

Développement des grilles

La phase de développement a débuté avec une analyse des besoins relativement au cours de *Mécanique pour ingénieurs* dans lequel s'est déroulé le projet et en parallèle, une recherche documentaire a été réalisée pour repérer des exemples de grilles pour la résolution de problèmes, v. par ex. [3, 8]. Cela a mené au constat qu'il n'y avait pas d'outils appropriés pour satisfaire aux besoins pédagogiques du cours. Entre autres choses, l'analyse du format et du contenu des examens a révélé qu'il était préférable de construire deux grilles générales, une pour la schématisation et une autre pour la résolution avec applications numé-

riques. Cette analyse a aussi signalé qu'il était plus pertinent d'élaborer des grilles à échelles descriptives du type global, car il était plus approprié d'exiger des correcteurs de porter un seul jugement global à chaque sous-question des problèmes. L'utilisation de grilles du type analytiques aurait été trop laborieuse. Les résultats de l'analyse de la situation ont conduit à l'élaboration de propositions de grilles, puis à un processus itératif entre leur mise à l'essai et leur perfectionnement. La Figure 1 illustre la grille développée pour la résolution de problèmes. Elle contient 6 échelons.

Echelons	Descripteurs
Excellent	La solution s'appuie sur les principes physiques appropriés. Elle renferme les équations de base ainsi que les informations (variables, contraintes, forces, etc.) suffisantes et nécessaires pour représenter la réalité physique du problème. La démarche est complète et juste. Bref, il démontre une maîtrise de la matière dans sa démarche.
Très bien	La solution s'appuie sur les principes physiques appropriés. Elle renferme les équations de base ainsi que les informations (variables, contraintes, forces, etc.) suffisantes et nécessaires pour représenter la réalité physique du problème. La solution comprend une ou deux erreurs mineures . La démarche est complète. Bref, l'étudiant démontre une excellente compréhension de la matière dans sa démarche et semble maîtriser tous les concepts importants.
Bien	La solution s'appuie sur les principes physiques appropriés. Elle renferme toutes les équations de base appropriées et celles-ci contiennent la plupart des éléments représentant la réalité physique du problème. La solution comprend quelques erreurs mineures . Au regard des équations et des informations retenues, la démarche est complète. Bref, il démontre une bonne compréhension de la matière dans sa démarche et semble maîtriser la plupart des concepts importants.
Acceptable	La solution s'appuie sur les principes physiques appropriés. Elle renferme certaines équations de base appropriées et celles-ci contiennent quelques informations représentant la réalité physique du problème. La solution comprend plusieurs erreurs mineures OU La solution comprend un erreur majeure . Au regard des équations et des informations retenues, la démarche est incomplète. Bref, il démontre une compréhension acceptable de la matière dans sa démarche et ne semble pas maîtriser certains concepts importants.
Insuffisant	La solution s'appuie en partie sur les principes physiques appropriés. Elle renferme quelques équations de base appropriées et celles-ci contiennent quelques informations représentant la réalité physique du problème. La démarche est incomplète . Bref, il démontre une compréhension rudimentaire de la matière et ne semble pas maîtriser plusieurs concepts importants.
Très insuffisant	La solution ne s'appuie pas sur les principes physiques généraux appropriés à la résolution du problème. Ou L'étudiant ne formule aucune solution Bref, il démontre une compréhension insatisfaisante de la matière.

Figure 1. Grille à échelle descriptive de résolution de problèmes

Expérimentation des méthodes de correction

Au moment où le format et le contenu des grilles sont devenus raisonnablement stables, la deuxième phase de la recherche a été entreprise. Les chercheurs ont invité quatre correcteurs expérimentés à participer à cette seconde phase. Ces correcteurs étaient tous des étudiants aux cycles supérieurs et ils avaient corrigé à quelques reprises des examens dans le cours de Mécanique pour ingénieurs.

Au départ, les correcteurs ont été appelés à corriger individuellement selon la méthode de correction classique les copies d'un échantillon de 30 étudiants (8 faibles, 14 moyens, 8 forts), et ce pour sept problèmes d'examen. La répartition des copies selon les aptitudes anticipées des étudiants a été effectuée en s'appuyant sur les résultats du test diagnostic Force Concept Inventory [9] que les étudiants devaient passer lors des deux premières semaines du cours.

Ces mêmes correcteurs ont ensuite été conviés à une séance de formation en groupe à l'utilisation des grilles à échelles descriptives, au cours de laquelle le modérateur leur a expliqué la démarche de correction et leur a présenté quelques copies d'examen pour une correction collective. Cela a permis aux correcteurs de mettre à l'essai les grilles et de partager leurs interprétations au sujet de leur correction.

Enfin, les correcteurs ont reçu un deuxième échantillon de 30 copies d'étudiants pour effectuer une correction avec les grilles. Les sept problèmes de la première étape ont été retenus pour cette étape. À la fin

de chaque étape de cette phase d'expérimentation, les correcteurs ont été invités à relater la perception de leur expérience de correction.

L'ANALYSE DES RESULTATS

Les analyses sont divisées en deux axes dans cette étude. Les données quantitatives manifestent les scores décernés par les correcteurs dans les deux méthodes de correction et les données qualitatives témoignent de la perception des correcteurs par rapport aux deux approches.

Les scores décernés lors des deux corrections

Afin de procéder à des analyses comparatives, les scores des deux méthodes de correction ont d'abord été normalisés. Dans la correction classique, les correcteurs devaient noter les démarches des étudiants sur des échelles variant entre 5 et 20 points. Dans la correction avec grilles, les correcteurs pouvaient noter les solutions entre les six échelons des grilles, mais aussi ils pouvaient accorder un demi-échelon.

La figure 2 indique les scores moyens en pourcentage décernés par chaque correcteur dans les deux méthodes de correction. Les correcteurs B et C ont maintenu des scores moyens similaires dans les deux approches, alors que les correcteurs A et C ont attribué des résultats légèrement inférieurs dans la méthode de correction avec les grilles.

Correcteur	A	B	C	D	Total
Méthode classique	56%	68%	60%	69%	63%
Méthode avec grilles	53%	69%	55%	69%	61%

Figure 2. Comparaison des moyennes des scores décernés dans les deux méthodes de correction

Les échelles des deux méthodes ne contiennent pas le même nombre de points. Même si certaines échelles dans la méthode classique comportaient jusqu'à 20 points, il a été observé que les correcteurs n'emploient qu'un nombre restreint d'échelons. La figure 3 compare le nombre moyen de points d'échelle dans les deux méthodes. Le correcteur A a recouru en moyenne à plus d'échelons lors de la correction avec grilles tandis que les correcteurs B et C ont utilisé moins d'échelons dans cette méthode de correction.

Correcteur	A	B	C	D
Méthode classique	5,7 (2,9)	7,5 (2,5)	6,8 (2,7)	5,8 (1,9)
Méthode avec grilles	6,3 (1,6)	5,9 (1,4)	6,1 (1,8)	5,8 (1,1)

Figure 3. Nombre moyen de points d'échelle et écart-type pour les deux méthodes de correction

Les données quantitatives récoltées lors de la correction classique et de la correction avec la grille ont aussi été analysées dans le cadre de la théorie de généralisabilité qui se veut le prolongement des mo-

dèles de la théorie classique des tests et qui s'inscrit dans la logique de l'analyse de variance [10, 11]. Cette forme de modélisation vise principalement à différencier et à quantifier les sources d'erreur, sous forme de proportion de la variance d'erreur absolue, dans une opération de collecte de données. Le diagramme d'Euler-Venn de la Figure 4 illustre le plan de l'étude de généralisabilité de cette étude pour chaque méthode de correction. Les étudiants (E) forment la facette de différenciation, alors que les correcteurs (C) et les items (I) constituent les facettes d'instrumentation qui contribuent à l'erreur de mesure. L'erreur due à l'instrumentation est en conséquence composée des effets principaux des correcteurs et des items, mais aussi des interactions entre les différentes facettes. Cette décomposition des sources de variance permet également de calculer un coefficient de généralisabilité pour décision absolue, noté ρ^2 , qui renvoie à la proportion de la variance attribuée à la facette de différenciation (E) par rapport à la variance totale. Ce coefficient est un indice de fidélité des résultats.

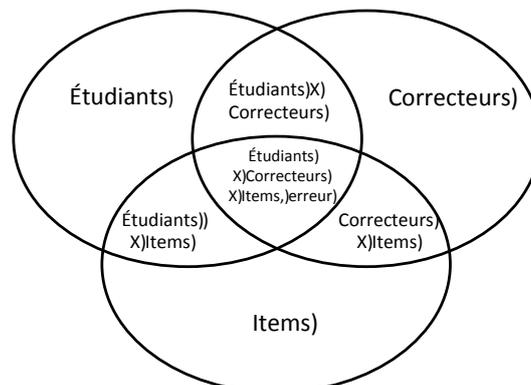


Figure 4. Diagramme d'Euler-Venn pour l'étude de généralisabilité Étudiants X Items X Correcteurs

La Figure 5 illustre la proportion de la variance de l'erreur absolue dans les deux méthodes de correction. Il faut d'abord constater que les effets C et I ainsi que l'interaction EI contribuent en grande partie à la variance totale de l'erreur. Ces résultats montrent que l'effet dû à la variance de la sévérité des correcteurs est presque deux fois plus important dans la correction avec la grille que dans la méthode de correction traditionnelle. L'effet des items concerne la variance attribuable au degré de difficulté des items, alors que l'interaction entre les étudiants et les items signifie que chaque étudiant perçoit différemment le degré de difficulté des items. La proportion élevée de la variance accordée à cette composante signale que les résultats des étudiants sont possiblement influencés par des éléments contextuels des problèmes. D'autres recherches ont constaté des observations similaires dans le cadre de tâches complexes en science [12]. L'étude de généralisabilité indique également un coefficient de généralisabilité d'une valeur de 0,67 pour la correction classique et d'une valeur de 0,59 pour la correction

avec grille. Cela suggère que la correction classique donne des résultats plus fidèles.

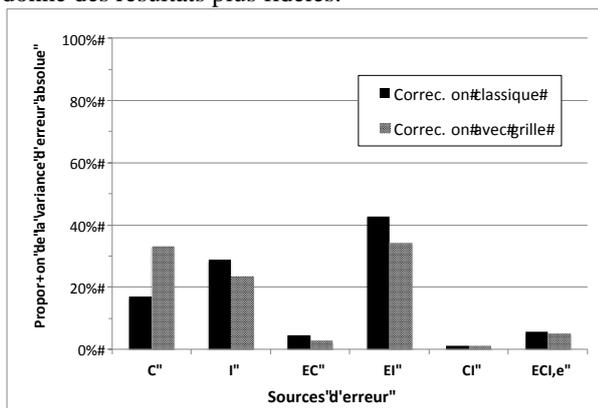


Figure 5. Comparaison de la répartition des sources d'erreurs pour les deux méthodes

Les observations de la séance de formation

Les échanges entre le modérateur et les correcteurs lors de la session de formation ont été enregistrés sur bande audio, puis les chercheurs ont procédé à une analyse du contenu de ces discussions. Voici les constats à la suite de cette analyse. Les correcteurs se sont dits à l'aise à utiliser les grilles à échelles descriptives, ils ont indiqué qu'il était suffisamment facile et rapide d'apprendre à corriger à l'aide des grilles et ils ont indiqué qu'ils étaient satisfaits de la rencontre de groupe. En contrepartie, l'entretien de groupe a révélé certaines difficultés en ce qui concerne l'utilisation des grilles. Les correcteurs ne détectent pas les mêmes erreurs dans les démarches des (mêmes) étudiants, ils n'accordent pas la même importance aux erreurs qu'ils détectent et ils n'adoptent pas le même raisonnement pour justifier le choix de l'échelon de la grille de correction.

Les entretiens individuels

Une série d'entretiens individuels avec les correcteurs après les deux blocs de correction a permis de recueillir leur perception au sujet des méthodes de correction. Trois des quatre correcteurs ont révélé qu'ils se sentaient plus à l'aise à travailler avec une grille de correction. Ils ont précisé qu'elle leur permettrait de corriger plus efficacement, et que cette méthode serait un atout pour atténuer les différences entre les correcteurs, mais aussi pour aider les nouveaux correcteurs. Par contre, ils ont mentionné que la grille réduirait la précision des scores et qu'elle entraînerait une diminution des résultats des étudiants aux examens. En effet, les correcteurs ont déclaré avoir l'impression que la grille rendait leur jugement plus sévère.

CONCLUSIONS

L'évaluation de la résolution de problèmes dans de grandes classes exige de mettre en place des mécanismes pour assurer une démarche juste et équitable pour tous les étudiants. D'abord, les résultats de cette étude montrent que les scores moyens des étudiants

dans les deux méthodes de correction ne semblent pas varier de façon importante, contrairement aux opinions des correcteurs. Il faut toutefois être prudent parce qu'il est possible qu'il existe une différence réelle entre les deux échantillons d'étudiants, ce qui n'a été contrôlé qu'indirectement dans la présente étude, en formant des regroupements d'étudiants sur la base des résultats au Force Concept Inventory.

Ensuite, certains correcteurs ont déclaré que la méthode de correction avec des grilles réduirait la précision des scores. Il faut cependant constater que le nombre de points d'une échelle ne représente pas la « précision », car il n'offre aucune information sur l'erreur de mesure. Il faut recourir à d'autres mécanismes pour assurer la fidélité des résultats. La décomposition de la variance dans l'étude de généralisabilité met en évidence l'effet des correcteurs dans les deux approches et notamment, elle attire l'attention sur la plus grande variabilité du degré de sévérité des correcteurs dans la correction avec grilles. L'analyse des échanges de la formation de groupe apporte néanmoins un éclairage essentiel quant à l'explication de ces différences entre les méthodes de correction. Il a été observé que les correcteurs ne repèrent pas les mêmes erreurs dans les démarches des étudiants, ils ne les interprètent pas de façon identique et ils n'adoptent pas les mêmes justifications pour expliquer les scores décernés. Ces observations sont pertinentes pour les deux méthodes de correction.

En définitive, il apparaît que la correction avec les grilles à échelles descriptives ne permet pas à elle seule de réduire la variabilité des interprétations entre les correcteurs. Cette étude a signalé l'importance d'accompagner l'implantation de tout outil d'évaluation d'une formation appropriée afin d'assurer une compréhension commune du processus de correction. Le recours aux grilles à échelles descriptives permet dans ces circonstances d'offrir une structure cohérente pour harmoniser les pratiques de correction, dans la mesure où les correcteurs reçoivent une formation adéquate.

Bibliographie

- [1] Conseil canadien des ingénieurs, *Normes et procédures d'agrément*, 2011, Bureau canadien d'agrément des programmes de génie.: Ottawa.
- [2] Adams, W.K. and C.E. Wieman, *Problem Solving Skill Evaluation Instrument --- Validation Studies*. AIP Conference Proceedings, 2007. **883**(1): p. 18-21.
- [3] Docktor, J.L., *Development and validation of a physics problem-solving assessment rubric*, 2009, University of Minnesota: United States -- Minnesota.
- [4] Jonassen, D.H., *Learning to solve problems: a handbook for designing problem-solving*

- learning environments* 2011, New York: Routledge.
- [5] Moskal, B.M., *Scoring Rubrics: What, When and How?* Practical Assessment, Research and Evaluation, 2000. **7**(3).
- [6] Durand, M.-J. and R. Chouinard, eds. *L'évaluation des apprentissages: De la planification à la communication des résultats*. 2012, Marcel Didier: Montréal.
- [7] Moskal, B.M. and J.A. Leydens, *Scoring Rubric Development: Validity and Reliability*. Practical Assessment, Research and Evaluation, 2000. **7**(10).
- [8] Baril, M., R. Desbiens, and É. Guay, *La méthode de résolution de problème comme instrument pour faire acquérir une compétence plutôt que des connaissances*, 2006: Québec.
- [9] Hestenes, D., M. Wells, and G. Swackhamer, *Force Concept Inventory*. *Physics Teacher*, 1992. **30**(3): p. 141-58.
- [10] Brennan, R.L., *Generalizability theory* 2001, New York: Springer.
- [11] Cardinet, J., S. Johnson, and G. Pini, *Applying Generalizability Theory using EduG* 2010, New York: Routledge.
- [12] Shavelson, R.J., G.P. Baxter, and X. Gao, *Sampling variability of performance assessments*. *Journal of Educational Measurement*, 1993. **30**(3): p. 215-232.