

Développement et validation d'un test mesurant la compréhension des concepts cinématiques en physique au secondaire

Louis Trudel, Carmen Parent and Réjean Auger

Volume 31, Number 1, 2008

URI: <https://id.erudit.org/iderudit/1025014ar>

DOI: <https://doi.org/10.7202/1025014ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Trudel, L., Parent, C. & Auger, R. (2008). Développement et validation d'un test mesurant la compréhension des concepts cinématiques en physique au secondaire. *Mesure et évaluation en éducation*, 31(1), 93-120.
<https://doi.org/10.7202/1025014ar>

Article abstract

In spite of the importance of understanding in science learning, there is a shortage of tests that measure the level of understanding of high school students in physics. Thus, the objective of our research is to develop a test measuring the students' level of understanding of kinematical concepts in high school physics which is valid as a diagnostic tool. The analysis of answers by an item response model shows that the levels of understanding defined by the test are well separated and constitute a hierarchical and cumulative scale.

Développement et validation d'un test mesurant la compréhension des concepts cinématiques en physique au secondaire

Louis Trudel

Université d'Ottawa

Carmen Parent

Université du Québec à Montréal

Réjean Auger

Professeur honoraire chez OgDDASt inc.

MOTS CLÉS: Mesure de la compréhension, cinématique, modèle de réponse à l'item, super-item, taxonomie SOLO

Malgré l'importance de la compréhension des concepts dans l'apprentissage des sciences, il existe peu de tests qui permettent de mesurer le niveau de compréhension des élèves. Par conséquent, l'objectif de cet article est de développer un test mesurant le niveau de compréhension des concepts cinématiques en physique au secondaire qui soit valide à des fins diagnostiques. L'analyse des réponses à l'aide d'un modèle de réponse à l'item montre que les niveaux de compréhension définis par le test sont distincts les uns des autres et forment une échelle hiérarchique et cumulative.

KEY WORDS: Comprehension assessment, kinematics, item response model, super-item, SOLO taxonomy, formative or diagnostic assessment

In spite of the importance of understanding in science learning, there is a shortage of tests that measure the level of understanding of high school students in physics. Thus, the objective of our research is to develop a test measuring the students' level of understanding of kinematical concepts in high school physics which is valid as a diagnostic tool. The analysis of answers by an item response model shows that the levels of understanding defined by the test are well separated and constitute a hierarchical and cumulative scale.

PALAVRAS-CHAVE: Avaliação da compreensão, cinemática, modelo de resposta ao item, super-item, taxonomia SOLO, avaliação formativa ou diagnóstica

Apesar da importância da compreensão dos conceitos na aprendizagem das ciências, existem poucos testes que possibilitem medir o nível de compreensão dos alunos. Assim, o objectivo deste artigo é o de desenvolver um teste que permita medir o nível de compreensão dos conceitos cinemáticos em Física, no ensino Secundário, que seja válido para elaborar diagnósticos. A análise das respostas, com a ajuda de um modelo de resposta ao item, mostra que os níveis de compreensão definidos pelo teste são distintos uns dos outros e formam uma escala hierárquica e cumulativa.

Introduction

Le domaine de la physique qui étudie le mouvement sans se préoccuper de ses causes s'appelle la cinématique. L'apprentissage des concepts cinématiques est important parce que ceux-ci forment la base de la mécanique classique (Halloun, 1996). Or, la cinématique est un domaine qui cause beaucoup de difficultés aux élèves, principalement parce que ceux-ci interprètent les phénomènes du mouvement à l'aide de schémas qui se sont élaborés progressivement lors d'interactions avec des phénomènes de leur quotidien (Arons, 1990). Ces schémas diffèrent de façon marquée des concepts scientifiques et, dans certains cas, peuvent même interférer avec l'apprentissage. Les incompréhensions qui en résultent ne sont pas habituellement détectées par les tests traditionnels. En effet, il semble que les élèves soient capables de maîtriser les méthodes de calcul sans adopter la conceptualisation qui les soutient, de sorte qu'un schéma erroné peut ne pas être détecté parce que la connaissance des formules et l'utilisation de techniques superficielles de résolution masquent l'incompréhension des concepts (Masters & Mislevy, 1993). Pour pallier cette situation, de nouvelles méthodes d'évaluation ont été proposées qui visent à évaluer la pensée de l'élève en action, par exemple lorsqu'il effectue une tâche ou résout un problème. Les explications fournies par l'élève ou les observations faites lorsqu'il effectue ces tâches sont enregistrées ou notées et, par la suite, analysées en vue de distinguer sa façon de comprendre (White & Gunstone, 1992). En utilisant cette approche, des recherches tant transversales que longitudinales révèlent une progression dans l'acquisition des concepts du mouvement chez les élèves, d'un niveau inférieur de compréhension,

où les élèves confondent la vitesse et l'accélération, aux niveaux supérieurs où ils démontrent une compréhension de l'accélération en tant que rapport entre le changement de vitesse et l'intervalle de temps correspondant (Bliss, Morrison & Ogborn, 1988 ; Trowbridge & McDermott, 1980, 1981).

Même si ces recherches ont fourni des renseignements valables sur la compréhension des élèves des phénomènes du mouvement et des concepts s'y rapportant, les méthodes utilisées sont peu pratiques dans un contexte de classe et l'interprétation de leurs résultats par l'enseignant est malaisée (Sadler, 2000). C'est pourquoi, dans un désir de systématiser la collecte des données et leur analyse afin d'établir un diagnostic rapide et efficace des carences dans la compréhension des élèves, des tests objectifs ont été réalisés en s'inspirant des recherches sur les schémas des élèves en cinématique. La procédure consiste en premier lieu à identifier les principaux schémas que les élèves possèdent concernant un concept cinématique, puis de concevoir des items à choix multiples où la réponse correcte est choisie dans un ensemble incluant des leurres, constitués des principaux schémas repérés chez les élèves (Treagust, 1995).

Élaboré d'après ces principes, le test d'Halloun et Hestenes (1985) vise à identifier les connaissances initiales des étudiants avant d'entreprendre des cours d'introduction à la physique, c'est-à-dire les connaissances dérivées de leurs expériences quotidiennes avec les phénomènes du mouvement. Quoique la fidélité du test soit excellente (le coefficient de fidélité de Kuder-Richardson se situe entre 0,86 et 0,89), les résultats obtenus par les étudiants à ce test avant un cours d'introduction à la physique se sont révélés similaires à ceux obtenus après le cours (Halloun & Hestenes, 1985). Même si Halloun et Hestenes (1985) expliquent cette situation par le caractère traditionnel de l'enseignement de la physique à l'université, il est possible que la relative insensibilité de ce test à l'enseignement reflète une carence du test à mesurer certains apprentissages fondamentaux en physique, tels que la résolution de problèmes. C'est une des conclusions à laquelle est arrivé Dickie (1988) qui, en utilisant une version adaptée de ce test à des étudiants anglophones du secteur collégial au Québec, a montré que les résultats de ce test ne permettaient pas de prévoir leur performance future au cours de physique. Par ailleurs, une analyse en composantes principales d'un test similaire (c.-à-d. la version élaborée par Hestenes, Wells et Swackhamer, 1992) a révélé que les problèmes de cette version ne sont que faiblement reliés et, par conséquent, ne mesurent pas seulement le concept de force ou ses principales dimensions tels que visés (Huffman & Heller, 1995). Selon ces derniers auteurs, le caractère fragmenté

des connaissances des étudiants pourrait expliquer ces résultats. À cet égard, il apparaît que le degré de structure des connaissances des élèves en mécanique se situe entre le caractère fragmenté rapporté précédemment et le caractère plus cohérent d'idées reliées ensemble par des relations de nature causale (Oliva, 1999).

En conclusion, malgré les multiples perspectives offertes par ces recherches concernant la compréhension du mouvement, certains constats émergent :

1. si certaines connaissances des élèves sont propres au contexte dans lequel elles se sont développées, d'autres sont reliées sous forme de cadres causaux de plus grande généralité ;
2. l'élève suit une série d'étapes lors de sa progression d'une compréhension naïve à une compréhension plus scientifique des phénomènes du mouvement.

Pour expliquer cette progression, certains auteurs ont avancé que l'interaction entre les structures mentales de l'apprenant et les caractéristiques de la tâche produit un ensemble de réponses qui diffèrent par leur degré de complexité (Bereiter, 2002 ; Biggs, 2003). En adoptant ce point de vue, il est possible de considérer la compréhension, non pas de façon ponctuelle, mais plutôt selon un axe de développement conceptuel qui permet de repérer les progrès de l'élève et de prescrire les interventions appropriées (Wilson, 2005). Eu égard à cette situation, nous poursuivons l'objectif de proposer un test qui soit valide pour diagnostiquer le niveau de compréhension des concepts cinématiques en physique au secondaire.

Méthode

Afin de mesurer le niveau de compréhension, Masters et Mislevy (1993) suggèrent, d'une part, que les items du test contiennent une série de questions destinées à révéler les différents niveaux de compréhension des élèves et, d'autre part, que ces niveaux de compréhension soient par la suite comparés à un cadre de référence qui permette de les ordonner. Un type d'items qui satisfait le premier critère de Master et Mislevy (1993) est ce que Collis et Davey (1986) désignent comme un super-item. Un super-item se construit en deux parties. La première partie, appelée tronc commun, est un énoncé décrivant une situation physique mettant en œuvre certaines propriétés choisies. La deuxième partie consiste en une série de questions dans le but d'identifier les différents niveaux de compréhension des élèves à propos de la situation physique décrite dans le tronc commun (Collis & Davey, 1986 ; Haladyna, 2004).

Puisque le présent test porte sur la cinématique, le tronc commun du super-item mettra en jeu les différents concepts du mouvement et leur organisation. En cinématique, il existe trois modèles principaux qui explicitent les simplifications, les liens, les contraintes et les structures des phénomènes : le mouvement rectiligne uniforme, le mouvement uniformément accéléré et le mouvement mixte (Halloun, 1996). Les situations physiques décrites dans le tronc commun du super-item doivent donc pouvoir être classées dans l'un ou l'autre de ces modèles (Trudel, 2005).

Pour satisfaire le second critère de Masters et Mislevy (1993), les questions du super-item sont formulées de façon à faire varier le degré de complexité du traitement de l'information exigé de l'élève pour les accomplir avec succès, en accord avec les catégories de la taxonomie SOLO proposée par Biggs et Collis (1982) : unistrukturé, multistrukturé, relationnel et étendu-abstrait.

Ces catégories sont définies de la façon suivante (Biggs et Collis, 1982) :

- Unistrukturé (U) : L'utilisation d'une information ou d'une donnée fournie de façon explicite dans l'énoncé.
- Multistrukturé (M) : L'utilisation de façon séquentielle de deux ou de plusieurs opérations discrètes basées sur l'information fournie dans l'énoncé pour formuler une conclusion unique.
- Relationnel (R) : L'intégration de l'information fournie dans l'énoncé pour formuler une conclusion unique ou une généralisation.
- Étendu-abstrait (ÉA) : L'utilisation de plusieurs systèmes abstraits interactifs pour formuler une réponse. Ceci pourrait inclure la formulation d'une hypothèse, l'évaluation de la qualité d'un modèle et la formulation d'une réponse élaborée. Ce niveau de réponse requiert l'utilisation de certaines données qui ne sont pas fournies dans l'énoncé.

Ainsi, le problème 5 (voir annexe) fournit un exemple de super-item dont le tronc commun est constitué d'un énoncé, d'une figure et d'un graphique qui communiquent respectivement certaines données sur le type de mouvement, la description du mouvement, et les différentes positions en fonction du temps d'une bille remontant un plan incliné. Les trois questions (5.1, 5.2 et 5.3) qui s'y rattachent exigent de l'élève de traiter l'information de l'énoncé successivement au niveau unistrukturé (distance demandée), au niveau multistrukturé (déplacement demandé) et au niveau relationnel (vitesse moyenne demandée).

Afin de déterminer si les séries de questions associées aux différents super-items forment une échelle hiérarchique et cumulative et de calibrer celle-ci, nous avons choisi le modèle de l'échelle d'appréciation (Andrich, 1988). Ce modèle a été préféré au modèle de crédit partiel parce que, comme nous le verrons au paragraphe suivant, les seuils de transition entre les niveaux de compréhension dans le modèle de l'échelle d'appréciation sont indépendants des contenus cinématiques des énoncés, ce qui n'est pas le cas dans le modèle de crédit partiel. Cette indépendance des seuils relativement aux contenus d'un ensemble d'items d'un domaine spécifique de connaissances est en accord avec les fondements théoriques de la taxonomie SOLO (Hattie & Purdie, 1998).

Dans le modèle de l'échelle d'appréciation, le calcul des probabilités de réussite s'obtient avec l'équation suivante (Bond & Fox, 2001) :

$$P_{nik} = \exp(\beta_n - \delta_{ik}) / [1 + \exp(\beta_n - \delta_{ik})],$$

où β_n est l'habileté de l'élève n en logits¹

$\delta_{ik} = (\delta_i + \tau_k)$ où δ_i est la difficulté de l'item i en logits et

τ_k est la mesure du seuil de réponse associé à chacun des sauts de compréhension « k ».

Ce modèle donne la probabilité « P_{nik} » qu'un élève d'un niveau de compréhension donné « β_n » puisse résoudre un super-item de degré de difficulté spécifique « δ_{ik} ».

Le double indice dans l'expression δ_{ik} sépare le degré de difficulté δ_{ik} associé au super-item en deux composantes : une sur la difficulté du super-item δ_i et l'autre sur la difficulté de chaque saut τ_k ($k = 1, 2, 3$). La valeur du paramètre associée à chaque saut, « τ_k », ne varie pas d'un super-item à l'autre. À partir de la réponse de l'élève, il est possible de situer sa compréhension sur une échelle pourvu que les différents super-items puissent être ordonnés suivant une ligne croissante de difficulté, en supposant l'unidimensionnalité de celle-ci (Bond & Fox, 2001).

Développement du test de compréhension des concepts cinématiques

En premier lieu, même si les questions des super-items peuvent être conçues sous un format ouvert, à réponse narrative (selon la formulation de Laveault et Grégoire, 2002), nous avons préféré les questions à choix multiple ou à réponse courte, car elles permettent de réduire la subjectivité de la correction et, par conséquent, augmente la fiabilité des résultats de la correction (Haladyna, 2004). Par ailleurs, certaines règles de rédaction doivent être

suivies pour faciliter les inférences que nous pourrions faire sur la compréhension à partir des résultats du test. Or, dans notre test, le rôle du tronc commun d'un super-item est double: d'une part, il doit mettre en œuvre les concepts du domaine choisi, soit la cinématique, et, d'autre part, il doit fournir suffisamment d'information sur la situation physique présentée pour que l'élève puisse répondre à différents niveaux de complexité selon la question posée concernant cette situation. Plus spécifiquement, la rédaction du tronc commun se fait en respectant les règles suivantes (Haladyna, 2004).

- Il doit porter sur les différents concepts de la cinématique, et en particulier les modèles de base de la cinématique décrits précédemment.
- Les situations décrites dans le tronc commun doivent concerner des phénomènes quotidiens. Nous avons choisi des situations de la vie quotidienne parce que ces situations sont familières aux élèves et que, par conséquent, elles sont plus susceptibles d'amener les élèves à exprimer leurs idées sur le mouvement qu'une situation abstraite ou scolaire.
- Au lieu de proposer aux élèves des problèmes dépouillés de tout contexte, nous leur avons présenté des situations physiques riches qui permettent à ces derniers de poser des jugements à divers niveaux de complexité (Biggs, 2003). Cette information riche est communiquée sous différentes formes: graphiques, tableaux ou dessins stroboscopiques. Ces divers modes de présentation permettent de communiquer un ensemble d'information sous une forme structurée et accessible qui permet à l'élève de choisir, parmi l'ensemble proposé, l'information qui est pertinente.

Les énoncés des troncs communs sont basés sur des idées et de l'information provenant de diverses sources. Nous nous sommes inspirés de livres d'enseignement de la physique utilisés au Québec en cinquième secondaire, de livres utilisés au premier cours de mécanique du collégial, de livres d'introduction à la physique provenant d'autres pays et de manuels didactiques sur l'enseignement de la physique (Trudel, 2005). En consultant ces différents ouvrages, nous avons établi une première liste de problèmes susceptibles de couvrir les différents concepts cinématiques visés par notre recherche: le déplacement, la vitesse et l'accélération des objets en mouvement. Par la suite, nous avons classé ces problèmes suivant les modèles physiques de base de la cinématique: mouvement rectiligne uniforme, uniformément accéléré et mixte. Ces opérations nous ont donné une panoplie de situations physiques pour lesquelles nous pouvions concevoir des questions associées aux divers niveaux SOLO.

Pour garantir le bon fonctionnement des questions et des réponses au sujet de la compréhension de l'élève, nous avons rédigé les questions des super-items selon les règles docimologiques suivantes (Haladyna, 2004).

- Afin de s'assurer que les questions construites correspondent à différents niveaux de la taxonomie SOLO, celles-ci doivent mettre en jeu les opérations que nous avons associées au processus de la compréhension : l'identification des facteurs pertinents du concept mesuré et la mise en relation de ces facteurs sous la forme d'une règle².
- Les questions des super-items sont formulées de façon concise sans que s'ajoute d'information nouvelle. Par conséquent, les principales caractéristiques de la situation étudiée sont exposées dans le tronc commun.
- Dans un super-item, il est important de garder les questions indépendantes les unes des autres. Pour assurer cette indépendance, nous avons construit, pour chaque super-item, une série de questions de sorte qu'il soit possible de répondre à chacune d'entre elles sans avoir répondu aux autres, c'est-à-dire que la réponse à une des questions n'est pas requise pour répondre aux autres. Cette caractéristique est importante pour assurer l'indépendance locale qui constitue, avec l'unidimensionnalité, les conditions d'utilisation du modèle de l'échelle d'appréciation.

La validation de cette version préliminaire a permis de constater que les résultats du test pouvaient être influencés par l'habileté à appliquer des formules (Collis & Davey, 1986). Pour diminuer l'influence possible de cette habileté sur les résultats du test, nous avons ajouté à notre test un second volet qui inclut des super-items mettant en jeu une compréhension de nature conceptuelle des différentes propriétés du mouvement. Nous avons pris soin de nous assurer que les deux volets de ce test suivent les règles docimologiques énoncées précédemment. Néanmoins, la rédaction des super-items appartenant à chacun des volets du test, le premier contenant des super-items plus associés à la résolution de problèmes et le second contenant des super-items de nature plutôt conceptuelle, diffère sous deux aspects principaux.

Ainsi, dans les questions portant sur les super-items appartenant au premier volet, par exemple le super-item 5 (voir annexe), la représentation des variables adopte une échelle proportionnelle et la relation exigée sur le plan relationnel s'exprime sous forme d'une règle quantitative, alors que dans les questions se rapportant aux super-items du deuxième volet, par exemple le super-item 12 (voir annexe), l'échelle utilisée est de nature ordinale et la relation exigée sur le plan relationnel constitue un enchaînement de proportionnalités qualitatives.

Validation du test de compréhension des concepts cinématiques

La décomposition d'un super-item en une partie fournissant l'information sur les propriétés du mouvement (tronc commun du super-item) et une partie constituée de questions à divers niveaux SOLO nous permet de vérifier respectivement la validité de contenu, qui vise à déterminer si le test porte bien sur les concepts cinématiques, et la validité conceptuelle, qui vise à déterminer si le test mesure bien le niveau de compréhension de l'élève (Collis & Davey, 1986). Pour vérifier l'atteinte de notre objectif de recherche, la validation de notre test doit répondre aux préoccupations suivantes (Trudel, 2005).

1. L'adéquation du test à la clientèle de cinquième secondaire et aux contraintes de l'enseignement au secondaire ou au contexte de la classe.
2. La mesure des niveaux de compréhension selon la taxonomie SOLO des propriétés du mouvement dans les principaux modèles de la cinématique.
3. La formation d'une échelle hiérarchique et cumulative, tel que prescrit par la taxonomie SOLO.

Validation de la version préliminaire

Dans un premier temps, seize super-items associés à la résolution de problème furent élaborés en utilisant quatre niveaux SOLO (unistructuré, multistructuré, relationnel et étendu-abstrait). Chaque super-item fut examiné par deux enseignants de physique de cinquième secondaire ayant chacun une classe d'élève dans cette discipline afin de vérifier si les concepts et les habiletés requises étaient appropriés aux élèves de cinquième secondaire. Par ailleurs, les élèves étaient également invités à faire des commentaires sur les difficultés reliées à la formulation des énoncés des super-items et des questions qui leur étaient associées. Leurs commentaires ont permis d'améliorer l'intelligibilité du test (Trudel, 2005). Cette étape a permis de s'assurer de la validité apparente du test, c'est-à-dire que le format du super-item est bien reçu des élèves et de leur enseignant et que le test peut être administré rapidement et efficacement dans un contexte de classe.

À la suite de cette première vérification, nous avons proposé la version préliminaire à trois enseignants, ayant chacun à leur charge un groupe d'élèves suivant le cours de physique 534. Au total, l'échantillon était composé de 85 élèves. Deux de ces enseignants œuvrent dans le secteur public, l'autre dans le secteur privé. Dans ce dernier cas, la classe est composée exclusivement de filles alors que les deux premières sont mixtes. Dans le but de vérifier

si les questions de chaque super-item formaient une échelle hiérarchique et cumulative, nous avons calculé, à partir des réponses fournies par ces élèves, les proportions de réponses correctes pour chaque niveau de compréhension de chaque super-item. Nous avons éliminé les super-items dans lesquels les niveaux SOLO étaient inversés. De même, nous avons vérifié jusqu'à quel point les réponses des élèves permettaient s'assigner sans ambiguïté une marque unique (Laveault & Grégoire, 2002). À cette fin, un coefficient de reproductibilité fut calculé. Les coefficients de reproductibilité des super-items retenus variaient de 0,8 à 1,0, ce qui est comparable au seuil de 0,85 adopté par Collis et Davey (1986). L'on peut donc dire que les questions des super-items retenus forment une échelle hiérarchique et cumulative au sens de Guttman (Stouffer, Guttman, Suchman, Lazarfeld, Star & Clauser, 1950).

La proportion de réussite à la question 4 (associée à un niveau étendu-abstrait de raisonnement) de tous les super-items est très basse. Cette faible proportion de réussite au niveau 4 concorde avec les résultats de Collis et Davey (1986), quant au faible nombre d'étudiants ayant atteint le stade formel de raisonnement. Or, puisqu'elle a été échouée dans une très large mesure, la question de niveau 4, de niveau étendu-abstrait, n'apporte que très peu d'information, à partir de l'échantillon d'élèves, sur la mesure de la compréhension. Par conséquent, la question 4 est ignorée dans le calcul du score total de l'ensemble des super-items du test. Mais dans une visée d'observation de la progression des élèves dans les divers niveaux de la compréhension, cette question est tout de même pertinente. Par ailleurs, un doute persiste quant à la validité conceptuelle du test construit uniquement à partir de questions associées à la résolution de problèmes qui requièrent l'utilisation de formules. En effet, l'étude des réponses des élèves aux questions des différents super-items SOLO a mis en évidence une influence possible de l'habileté à appliquer des formules dans l'obtention des réponses aux questions. Pour vérifier l'impact de cette habileté sur la validité de construit, nous avons décidé d'adjoindre un second volet au test de compréhension qui mesure la compréhension de nature conceptuelle des propriétés du mouvement.

Ajout d'un volet conceptuel au test de compréhension

À la lumière des résultats de la validation préliminaire, un second volet du test a été élaboré composé de huit super-items faisant appel à un raisonnement de nature conceptuelle. Ce volet a été soumis à l'approbation d'un didacticien des sciences afin de déterminer si les super-items le composant étaient appropriés aux élèves du secondaire et si ces super-items suivaient les mêmes règles docimologiques précédentes (Haladyna, 2004). Cette démarche a permis de modifier le contexte des situations physiques présentées à l'élève de façon à ce

qu'il se rapproche davantage de la vie quotidienne tout en mesurant les concepts du mouvement. En outre, ce type de situation contient des indices susceptibles d'activer les représentations que l'élève utilise pour interpréter le mouvement. Par la suite, les huit super-items le constituant ont été joints aux huit super-items du volet associés à la résolution de problème pour constituer la version de mise à l'essai du test de compréhension. Comme nous avons décidé d'enlever dans le calcul du score total le niveau étendu-abstrait de chaque super-item, ces seize super-items ainsi que leurs questions sous-jacentes permettent de couvrir les trois niveaux SOLO suivants : unistructuré, multistructuré et relationnel. Cette version a été soumise à une validation de contenu et de construit.

Le tableau 1 présente le tableau de spécification du test mesurant les divers niveaux de compréhension des concepts cinématiques en physique au secondaire dans sa version de mise à l'essai. L'ensemble du test présente deux volets (super-items de nature conceptuelle, et super-items associés à la résolution de problèmes).

Tableau 1
**Tableau de spécification de la version de mise à l'essai
du test de compréhension**

<i>Super-items Niveau SOLO</i>	<i>Super-item / résolution de problèmes</i>	<i>Super-item / nature conceptuelle</i>	<i>Total / Pourcentage relatif (%)</i>
U Unistructuré	1,1; 2,1; 4,1; 5,1 6,1; 7,1; 8,1; 9,1	3,1; 10,1; 11,1; 12,1; 13,1; 14,1; 15,1; 16,1	16 (33 1/3%)
M Multistructuré	1,2; 2,2; 4,2; 5,2; 6,2; 7,2; 8,2; 9,2	3,2; 10,2; 11,2; 12,2; 13,2; 14,2; 15,2; 16,2	16 (33 1/3%)
R Relationnel	1,3; 2,3; 4,3; 5,3; 6,3; 7,3; 8,3; 9,3	3,3; 10,3; 11,3; 12,3; 13,3; 14,3; 15,3; 16,3	16 (33 1/3%)
	24 questions 50%	24 questions 50%	48 questions 100%

Validité de contenu de la version de mise à l'essai du test

Puisque notre test porte sur la compréhension du mouvement, il est important de vérifier si les super-items portent bien sur les divers aspects du mouvement qui font l'objet de notre expérimentation. Il nous faut donc vérifier que chaque super-item du test peut être classé dans l'un ou l'autre des modèles de base de la cinématique décrits précédemment. Le procédé de validation utilisé consiste à soumettre le test à l'appréciation de plusieurs juges afin de déterminer le degré avec lequel ils s'accordent pour classer un super-

item spécifique dans l'un ou l'autre des modèles cinématiques (Laveault & Grégoire, 2002). À cette fin, le test a été soumis au jugement de dix étudiants qui suivaient un programme de formation des maîtres de première année en didactique des sciences ou en didactique de la physique. En effet, leur statut d'étudiant en formation des maîtres pour l'enseignement de la physique ou des sciences leur confère, selon nous, une certaine expertise en ce qui concerne la validation de contenu des super-items. D'après la valeur du coefficient Kappa (qui est de 0,61), les juges démontrent en général un degré d'accord qui est bon et significatif ($p < 0,001$) dans le classement de la majorité des super-items (Siegel & Castellan, 1988). Par conséquent, nous concluons que les super-items du test de compréhension portent bien sur les principaux modèles cinématiques, et ce, sur la base des jugements des étudiants maîtres en enseignement de la physique ou des sciences.

Validation conceptuelle de la version de mise à l'essai

Le test a été validé auprès d'un échantillon de 165 élèves composé de 82 garçons et de 83 filles. Cet échantillon de convenance provenait de six classes d'élèves étudiant la physique 534. Les enseignants de chacune des classes ont fait passer le test à leurs élèves. Cinq classes provenaient de la même école secondaire dans la région de Québec. Une classe provenait d'une école secondaire de la région métropolitaine de Montréal. Dans les deux cas, les élèves provenaient de milieux socio-économiques moyens et aisés. Les élèves qui composaient les groupes étaient considérés par leurs enseignants comme des élèves ayant de bonnes habiletés en sciences. Dans les cinq premières classes de l'école de la région de Québec, les élèves ont disposé d'une période de cours, soit soixante-quinze minutes, pour répondre aux questions du test. Dans la classe de la région de Montréal, les élèves ont bénéficié d'une période de soixante minutes seulement³ pour passer le test. Dans les deux cas, les élèves des six classes ont passé le test vers la fin de l'année, après qu'ils eurent terminé la partie de la cinématique du cours de physique 534.

Tous les élèves ont répondu à la même version du test. Un guide d'introduction accompagnait le test et s'adressait aux élèves, tandis qu'un guide à l'intention des enseignants contenait notamment des directives sur l'administration du test et l'envoi des résultats au chercheur. Ces instructions avaient pour but d'assurer l'uniformité des procédures de passation. Le test a été corrigé par les élèves avec un crayon d'une couleur différente. Les élèves ont été avertis que le test ne comptait pas, mais qu'il constituait une excellente préparation à un examen éventuel. Sitôt le test terminé et corrigé par les élèves, les enseignants nous ont envoyé les copies par la poste (Trudel, 2005).

Interprétation des résultats

Équivalence du volet conceptuel et du volet de résolution de problèmes

Dans la figure 1, nous avons comparé les valeurs des estimations des élèves de l'échantillon de validation concernant respectivement les deux volets du test. Si ces deux facettes mesurent le même trait latent, soit la compréhension des concepts de la cinématique, alors la valeur obtenue au sous-test (A) contenant les questions de nature conceptuelle devrait être sensiblement la même que la valeur obtenue au sous-test (B) associé à la résolution de problème eu égard aux erreurs de mesure. Ces mesures déterminent un intervalle de confiance. Le principe est que si on porte la valeur obtenue par l'élève au sous-test (A) en abscisse et si on porte la valeur de la compréhension au sous-test (B) en ordonnée, on devrait obtenir une droite à 45 degrés (c.-à-d. estimé (A) = estimé (B)) pourvu qu'une correction soit apportée pour éliminer la différence de difficulté entre les deux tests (Bond & Fox, 2001).

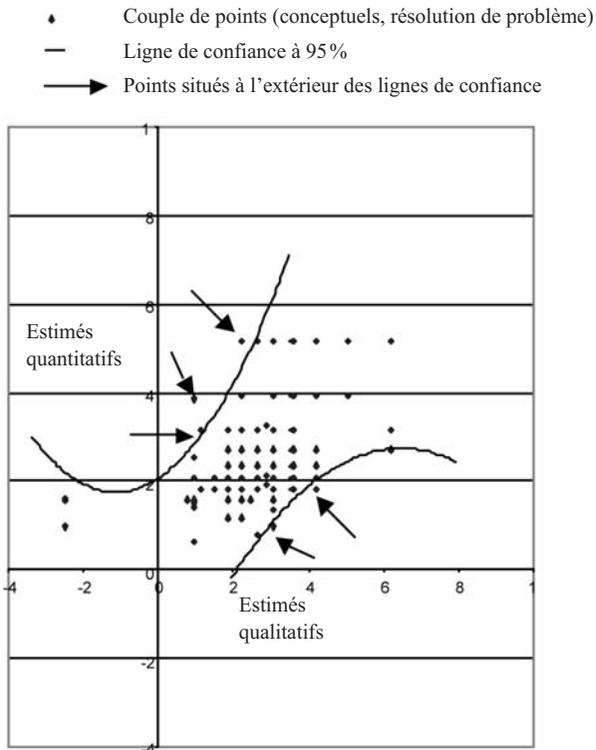


Figure 1. *Comparaison des estimés qualitatifs et quantitatifs*

L'intervalle de confiance est égal à deux fois l'erreur type moyenne obtenue en prenant la moyenne des erreurs type des deux estimés (Wright & Stone, 1979). Cet intervalle de confiance est tracé sur une droite perpendiculaire à la bissectrice du premier quadrant et passant par le point déterminé par les deux estimés. En reliant les points appartenant aux extrémités respectives des intervalles de confiance, nous obtenons deux courbes appelées ligne de confiance (voir figure 1). Si la majorité des points correspondant aux couples des estimés (A, B) se situent à l'intérieur de ces lignes de confiance, nous pouvons avancer que les estimés du volet conceptuel et les estimés du volet de résolution de problèmes ne se distinguent pas de façon significative les uns des autres. Ce qui est le cas présentement puisque seulement cinq couples de valeurs sur une possibilité de 165 sont à l'extérieur des lignes de confiance (indiqués par des flèches sur la figure 1), ce qui est au-dessus du seuil de confiance de 95%. En conclusion, nous constatons que les deux volets sont équivalents et que, par conséquent, nous pouvons les joindre afin d'en déterminer la validité de construit du test dans son ensemble. Par conséquent, les résultats obtenus par les élèves à ce test résulteront des contributions des super-items appartenant aux deux volets.

Propriétés métriques des super-items

Pour obtenir les propriétés des super-items, nous avons soumis les résultats de notre test à l'analyse du logiciel FACETS (Linacre & Wright, 1993). Ce logiciel analyse les réponses des élèves selon le modèle de l'échelle d'appréciation. Le tableau 2 donne le résultat de ces analyses. D'après le tableau 2, nous constatons tout d'abord que les super-items 2, 7, 8 et 15 ne se conforment pas au modèle de l'échelle d'appréciation car ils montrent des valeurs d'infit⁴ de -5, +5, +3 et -3 respectivement (cellules en gris dans le tableau 2). Ces problèmes, ayant un infit supérieur à +2 ou inférieur à -2, seront analysés à la section suivante pour tenter d'identifier les raisons de leur inadéquation avec le modèle de l'échelle d'appréciation et prendre éventuellement une décision éclairée sur leur maintien ou non dans le test en fonction de notre objectif de recherche. Pour situer les valeurs des degrés de difficulté des super-items et les degrés d'habiletés des élèves sur l'échelle de mesure, nous avons choisi l'origine de notre échelle de mesure à la valeur moyenne du degré d'habileté des élèves⁵. Nous constatons alors que la majorité des mesures des difficultés des super-items sont regroupées dans une fourchette allant de -3,13 logits à -1,05 logit, à l'exception des problèmes 1 (-4,82 logits) et 10 (-4,28) dont les valeurs sont sensiblement inférieures à la

moyenne et le problème 15 (0,69 logit) dont la valeur est sensiblement supérieure à la moyenne. Ces valeurs négatives des degrés de difficulté des super-items s'expliquent par le fait qu'ils sont généralement inférieurs aux degrés d'habiletés des élèves, compte tenu du choix de l'origine de notre échelle de mesure. Par ailleurs, l'erreur type de la mesure va de 0,12 logit à 0,17 logit, à l'exception du problème 10 (0,26 logit) et le problème 1 (0,30 logit).

Pour déterminer si les super-items sont suffisamment répartis en niveaux distincts pour définir une ligne d'intensité croissante de la maîtrise des concepts cinématiques, nous devons déterminer le degré de séparation de ceux-ci. Le degré de séparation de notre test se calcule à l'aide de la formule suivante :

$$G = SA / SE$$

où SA est l'écart type des degrés de difficulté des différents super-items du test ajusté en fonction de l'erreur de calibration et SE est la racine carrée de la moyenne des carrés des erreurs de calibration des super-items du test (Wright & Masters, 1982). D'après le logiciel FACETS, SA = 1,18 et SE = 0,16.

Par conséquent, le degré de séparation est $G = 1,18 / 0,16 = 7,31$. Ceci signifie que l'écart type ajusté des degrés de difficulté est 7,31 fois plus grand que la racine carrée de la moyenne des carrés des erreurs de calibration. Ce résultat indique un bon degré de séparation (Wright & Masters, 1982). La fidélité avec laquelle l'échantillon sépare les super-items est donnée par FACETS, soit 0,98. Par conséquent, nous concluons que les super-items sont suffisamment séparés pour définir une intensité croissante de la maîtrise des concepts cinématiques et que cette séparation est très fidèle. Par ailleurs, le nombre de strates distinctes dans lesquelles peuvent être regroupés les différents super-items du test est obtenu par la formule $(4G + 1)/3$, où G est le degré de séparation (Wright & Masters, 1982). Remplaçant la valeur de $G = 7,31$ dans la dernière équation, nous trouvons que les super-items peuvent être répartis en 10,08 strates distinctes. Par conséquent, le test permet de séparer les super-items en au moins dix strates distinctes quant au degré de difficulté des concepts cinématiques qui y sont proposés.

Tableau 2
*Paramètres des super-items lorsque la distribution des élèves
est centrée à l'origine*

<i>Super-item</i>	<i>Mesure de la compréhension (logits)</i>	<i>Erreur type</i>	<i>Infit</i>
1	-4,82	0,3	1
2	-2,43	0,13	-5
3	-2,91	0,14	1
4	-2,67	0,14	0
5	-3,13	0,16	0
6	-2,22	0,13	-1
7	-2,22	0,13	5
8	-2,64	0,14	3
9	-2,25	0,13	0
10	-4,28	0,26	0
11	-2,49	0,14	0
12	-1,91	0,12	-1
13	-2,1	0,13	0
14	-3,19	0,17	0
15	0,69	0,14	-3
16	-1,05	0,12	0
Test complet			
Moyenne	-2,48	0,15	0
Écart type	1,19	0,05	2,6

En ce qui concerne les degrés d'habileté des élèves, les valeurs de la variable latente sont comprises dans un intervalle variant de $-2,35$ à $+2,99$ (logits). La valeur ajustée de l'écart type des habiletés est $0,69$ et la racine carrée de la moyenne des carrés des erreurs de calibration est $0,50$. Par conséquent, le degré de séparation des degrés d'habiletés des élèves est $0,69 / 0,50 = 1,38$, avec une fidélité de $0,65$. Ce résultat est inférieur au degré de séparation des super-items. En calculant le nombre de strates distinctes de degrés d'habiletés avec la formule: $(4G + 1)/3$, nous obtenons un nombre de strates distinctes de $2,17$ (Wright & Masters, 1982). En conséquence, le test permet de séparer les élèves en deux groupes, ceux qui maîtrisent et ceux qui ne maîtrisent pas les concepts cinématiques du test, avec une fidélité moyenne (Trudel, 2005).

Analyse des super-items défectueux

Dans la section précédente, nous avons constaté que les super-items 2, 7, 8 et 15 ne se conforment pas au modèle de l'échelle d'appréciation. Or, l'utilisation de super-items en inadéquation avec le modèle de l'échelle d'appréciation peut nuire à la qualité des mesures effectuées par le test de compréhension. Il est donc important d'analyser les raisons des inadéquations et, à la suite de cette analyse, prendre la décision de conserver, modifier ou rejeter les super-items invalides sur une base empirique. Pour étudier les caractéristiques du super-item qui ont mené au désaccord, Wilson (1989) recommande une inspection des configurations de réponse en ayant à l'œil les résidus standardisés correspondants. Pour étudier la plausibilité des configurations de réponses, on compare, pour un même élève, les résultats obtenus au super-item défectueux aux résultats obtenus à des super-items de difficultés similaires, plus grandes ou plus petites. Une telle analyse nous a permis de constater que les super-items 2, 7, 8 et 15 ne se conforment pas à nos attentes. Dans le cas des super-items 7 et 8, certains niveaux sont inversés de sorte que l'échelle de compréhension obtenue n'est pas hiérarchique. Dans le cas des super-items 2 et 15, leur degré d'adéquation avec le modèle de l'échelle d'appréciation est trivial, car seuls certaines configurations de réponse sont permis, ce qui leur confère un faible pouvoir de discrimination.

Néanmoins, le retrait des super-items fautifs (2, 7, 8 et 15) du test de compréhension n'est pas toujours la meilleure solution, principalement à cause de l'information que ceux-ci apportent à l'ensemble du test (Linacre, 2002). En effet, le retrait éventuel de ces super-items aurait comme conséquence d'abaisser la fidélité et le degré de séparation de la mesure des habiletés des élèves (de 0,65 à 0,53 et de 1,38 à 1,06 respectivement) et des degrés de difficulté des super-items (de 0,98 à 0,97 et de 7,31 à 5,89 respectivement), ainsi que le nombre de strates distinctes dans les degrés de difficulté des super-items, qui passerait alors de 10 à 8. Ces changements auraient comme conséquence de diminuer la précision du diagnostic du degré de maîtrise des concepts cinématiques du test chez les élèves. Or, dans ce cas-ci, il semble que l'inadéquation de ces super-items soit en partie causée par des fréquences très faibles d'occupation de certains niveaux de compréhension, en particulier les niveaux préstructurés (catégorie 0) et unistructurés (catégorie 1). Cette situation peut s'expliquer par les caractéristiques de notre échantillon, constitué d'élèves ayant déjà terminé leur cours de physique de cinquième secondaire. Par conséquent, il est préférable, selon nous, de conserver ces super-items afin de préserver le caractère diagnostique associé au contenu de la cinématique et éventuellement de poursuivre la validation de cet instrument avec des échantillons d'élèves aux caractéristiques diversifiées.

Comparaison entre les degrés de difficultés et les habiletés des élèves

Nous présentons, à la figure 2, les distributions des habiletés des élèves et des degrés de difficulté des super-items. Il est à noter que les éléments ayant des valeurs extrêmes ne sont pas placés selon leur mesure, mais sont placés aux extrémités supérieures ou inférieures selon le cas dans la colonne appartenant à leur facette. Par ailleurs, le «+» ou le «-» indiquent que les mesures de la facette sont orientées positivement ou négativement. Enfin, dans la colonne de compréhension de l'élève, le symbole «*» représente la valeur «2» et le symbole «•» représente la valeur «1».

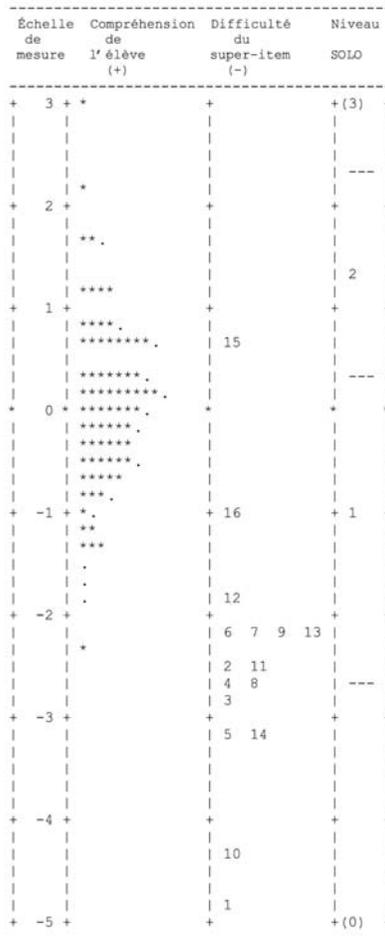


Figure 2. *Distribution comparée des habiletés des élèves, des degrés de difficulté et des niveaux SOLO*

La figure 2 nous indique que la majorité des degrés de difficulté se situe sous la distribution des habiletés des élèves, à environ 2,5 logits sous la valeur moyenne des habiletés. Deux super-items, le 1 et le 10, sont situés à 4 logits sous la moyenne de la distribution des habiletés. En ce qui concerne, les niveaux SOLO, leur position sur l'échelle de mesure de la figure 2 est déterminée en prenant la moyenne des habiletés des élèves qui ont répondu à un niveau SOLO, quel que soit le super-item du test. Sur cette échelle, les barres entre les niveaux indiquent les zones de transition entre les niveaux SOLO.

Propriétés de l'échelle de compréhension

Dans le modèle de l'échelle d'appréciation, l'étude des propriétés de l'échelle de compréhension nous renseigne sur la difficulté associée à chaque transition entre les niveaux SOLO et, en particulier, sur la probabilité d'atteindre un score particulier, les possibilités allant de 0 à 3, étant donné le niveau de compréhension de l'élève (β_n) et le degré de difficulté du super-item δ_{ik} ($\delta_{ik} = \delta_i + \tau_k$), associé au super-item δ_i et au saut τ_k . Le logiciel FACETS (Linacre & Wright, 1993) permet de calculer les difficultés des différents sauts entre les niveaux SOLO pour l'ensemble du test. Ces valeurs sont présentées dans le tableau 3 selon le score obtenu.

Tableau 3

Description des caractéristiques des divers sauts entre les niveaux SOLO

Score	Fréquence absolue	Fréquence relative (%)	Fréquence cumulée (%)	Saut	Difficulté du saut (logit)	Erreur type (logit)
0	22	1	1			
1	310	12	13	1	-2,61	0,23
2	721	29	42	2	0,69	0,07
3	1 450	58	100	3	1,92	0,05
Total	2 503	100				

Les catégories de la première colonne correspondent à l'ordre des questions de chaque super-item: préstructuré⁶ (0), unistructuré (1), multistructuré (2) et relationnel (3). La fréquence absolue d'utilisation de chaque catégorie apparaît dans la deuxième colonne. La fréquence relative en pourcentage apparaît dans la troisième colonne et la fréquence relative cumulée dans la quatrième colonne. Ces statistiques descriptives nous permettent d'apprécier l'importance des différentes catégories. Les trois dernières colonnes indiquent la position des sauts entre les niveaux SOLO ainsi que leurs erreurs types respectives (voir tableau 3).

Puisque 165 élèves ont rempli le questionnaire qui contient 16 super-items de trois questions chacun, une possibilité de $165 \times 16 = 2\,640$ scores possibles est à répartir entre les catégories 0, 1, 2 et 3. Le total des fréquences absolues indique 2 503 parce que certains élèves n'ont pas répondu à certaines questions. De ce total, 22 correspondaient à un score de 0, soit $(22 / 2\,503) \times 100 = 1\%$. La fréquence relative de cette catégorie est indiquée à la troisième colonne, soit 1% (tableau 3). En étudiant le tableau 3, on constate que la grande majorité des résultats aux différents super-items ont produit des scores de 3, soit $(1\,450 / 2\,503) \times 100 = 58\%$. Ce résultat est en accord avec le fait que les degrés de difficultés des super-items étaient inférieurs aux niveaux d'habiletés des élèves (voir graphique de la figure 2). Par ailleurs, en étudiant les trois dernières colonnes du tableau 3, on constate que les degrés de difficulté des différents sauts sont ordonnés selon les prescriptions de la taxonomie SOLO. Néanmoins, la position de ces différents sauts ne doit pas être confondue avec celle des niveaux de compréhension. En effet, la probabilité de réussite d'un élève à une question spécifique d'un super-item particulier dépend non seulement du degré de difficulté de cette question mais aussi du degré d'habileté de l'élève à répondre au test dans son ensemble. Par conséquent, plutôt que d'être associés à des positions précises sur l'échelle de mesure, les différents niveaux de compréhension sont définis sur des intervalles selon la probabilité qu'a l'élève de répondre à un niveau particulier plutôt qu'à un autre.

Pour déterminer ces intervalles, il nous faut tout d'abord composer une courbe de probabilité qui nous permet de trouver la probabilité d'atteindre les divers niveaux SOLO selon l'écart entre l'habileté de l'élève et le problème qui lui est soumis. Ainsi, dans la figure 3, chaque courbe donne la probabilité de réussir le saut correspondant, du niveau préstructuré (0) au niveau unistrukturé (1), du niveau unistrukturé (1) au niveau multistrukturé (2), du niveau multistrukturé (2) au niveau relationnel (3).

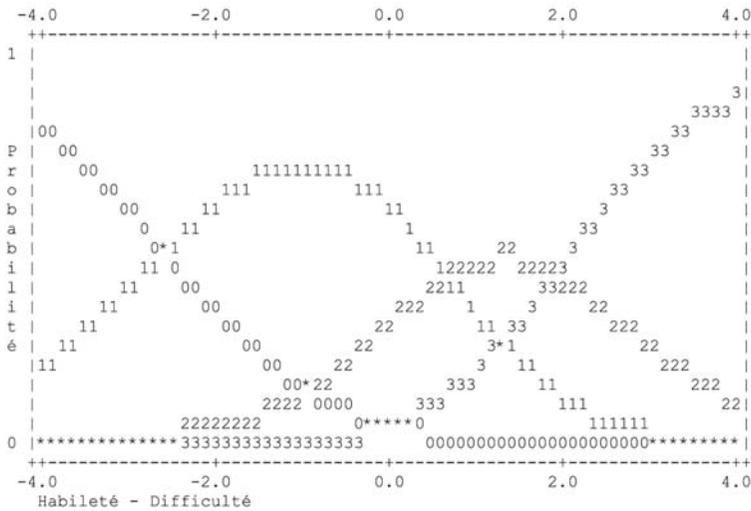


Figure 3. *Courbes de probabilité d'atteinte des divers niveaux SOLO en fonction de l'écart entre l'habileté de l'élève et la difficulté du niveau correspondant*

En étudiant la figure 3, nous voyons tout d'abord que les transitions entre les niveaux se font aux positions indiquées dans le tableau 3. Les points de transition sont déterminés par la plus petite mesure à laquelle il est le plus probable d'observer la catégorie. Par exemple, la transition du niveau préstructuré (0) au niveau unistructuré (1) se fait lorsque l'écart entre l'habileté de l'élève et la difficulté du super-item est supérieur à $-2,61$. De même, la transition du niveau 1 au niveau 2 se situe à $0,69$ d'écart et la transition du niveau 2 au niveau 3 s'opère à $1,92$ d'écart. Ces points de transition déterminent les intervalles dans lesquels la probabilité d'observer un niveau spécifique est supérieure à la probabilité d'observer les autres niveaux (voir figure 3). Par conséquent, à partir de ces points de transition, nous constatons que l'intervalle correspondant au niveau préstructuré est situé en dessous du point $-2,61$. Le niveau unistructuré est observé du point $-2,60$ au point $0,69$. Le niveau multistructuré s'étend du point $0,70$ au point $1,92$. Enfin, le niveau relationnel se situe au-delà du point $1,92$ (voir figure 3). En comparant les positions des différents niveaux sur l'échelle de «l'habileté – difficulté», nous constatons que les niveaux SOLO sont ordonnés tels que prévu et que, par conséquent, ils forment une échelle hiérarchique. En outre, les différents points de transition sont distincts statistiquement de ceux qui les suivent ou de ceux qui les précèdent sur l'échelle. En effet, le point de transition 1 de

mesure $-2,61$ a une erreur type de $0,23$. Or, le point de transition suivant ($0,69$) est situé au-delà de deux erreurs types du premier (voir tableau 3). En effet, $0,69 > -2,61 + 2 \times 0,23$. De même, le point de transition 2 ($0,69$) est déterminé avec une erreur type de $0,07$ (voir tableau 3). Ce point est statistiquement différent du point suivant. En effet, $1,92 > 0,69 + 2 \times 0,07$. Par conséquent, les différents niveaux de notre échelle de compréhension sont significativement distincts les uns des autres au seuil alpha de $0,05$.

Conclusion

L'importance de la compréhension dans l'apprentissage des sciences et la difficulté qu'éprouvent les méthodes traditionnelles à la mesurer ont amené les chercheurs à élaborer de nouvelles méthodes de l'évaluer. Le test de compréhension proposé dans cette recherche tente de combler ces lacunes en alliant la simplicité d'utilisation et la finesse de l'interprétation. Tout d'abord, notre test répond bien à la première exigence puisqu'il peut être administré dans le cadre d'une période de cours, qu'il est bien accepté par les enseignants et leurs élèves, qu'il est facile d'emploi et que son interprétation est aisée car les réponses des élèves permettent de situer leur compréhension en fonction des divers niveaux de la taxonomie SOLO. À cette fin, on peut aisément, à l'aide du graphique de la figure 3, déterminer la probabilité de réussite de l'élève à un super-item particulier (ou au test dans son ensemble), par suite du calcul de la différence entre le degré d'habileté de l'élève et le degré de difficulté d'un super-item particulier, et ainsi déterminer le niveau de compréhension de l'élève relativement au super-item choisi (ou au test dans son ensemble: voir section précédente). En ce qui concerne la seconde exigence, les énoncés des différents super-items décrivent bien des situations physiques mettant en œuvre les concepts clés du mouvement et leur organisation sous forme des principaux modèles de la cinématique. En outre, les super-items du test sont suffisamment séparés pour définir, avec une excellente fidélité, une ligne d'intensité croissante de la compréhension. Par contre, les élèves ne sont pas suffisamment séparés en groupes distincts quant à leur compréhension. Ce dernier constat s'explique peut-être par les caractéristiques de notre échantillon, constitué d'élèves ayant de bonnes habiletés en physique et ayant déjà terminé en classe la section sur la cinématique. Par ailleurs, les questions du niveau 4 dans la taxonomie SOLO sont actuellement ignorées dans le score total. Introduire celles-ci en y ajoutant des élèves dont les habiletés seraient au niveau formel de la compréhension permettrait d'augmenter l'étendue des niveaux taxonomiques du test et la validité de celui-ci.

Par ailleurs, les questions des super-items du test forment bien une échelle hiérarchique et cumulative. En effet, les niveaux de compréhension sont ordonnés suivant un degré de difficulté croissant, en accord avec les exigences de la taxonomie SOLO. En outre, les niveaux de compréhension sont significativement distincts les uns des autres de sorte qu'il est possible de les associer à des états de compréhension distincts. Toutefois, les valeurs obtenues de compréhension de l'élève sont moins précises que les valeurs de difficulté des super-items. Ceci s'explique par le fait que le degré de difficulté moyen de l'ensemble des super-items du test est situé sous le degré moyen des habiletés des élèves, de sorte que l'échelle que ces super-items définissent ne permet pas de cibler avec précision les divers degrés d'habileté des élèves après qu'ils aient suivi un cours de physique.

Pour augmenter la précision des mesures de la compréhension et assurer leur fidélité, il faudrait que les différents super-items soient davantage répartis le long de l'échelle de mesure. Une solution serait d'ajouter davantage de niveaux à l'échelle de compréhension, comme le niveau 4 que l'on n'a pas pris en considération. Également, plutôt que de considérer uniquement le développement dans le mode concret, le test pourrait également inclure des cycles de développement dans d'autres modes de représentation selon la progression suivante : iconique, symbolique, opératoire-concret et formel (Biggs, 2003).

Par ailleurs, l'équivalence entre les volets conceptuels et de résolution de problèmes semble surprenante, étant donné l'abondante documentation qui distingue entre la compréhension conceptuelle (Trowbrige & McDermott, 1981 ; Legendre, 2002) et la compréhension associée davantage à la capacité de résoudre des problèmes (Collis & Davey, 1986). Cette équivalence pourrait s'expliquer par des différences mineures de format entre les deux volets du test. Des recherches sont nécessaires afin de corroborer ce résultat.

Enfin, le fait que les points de transition entre les niveaux de compréhension soient ordonnés et distincts les uns des autres nous amène à conclure que le présent test peut servir des fins formatives. Ainsi, en identifiant le niveau de compréhension de l'élève à l'aide du test, l'enseignant pourrait, en fonction de ce diagnostic, planifier les activités à venir ou prescrire les interventions correctives appropriées (Biggs, 2003 ; Collis & Biggs, 1991). Par exemple, pour faciliter la transition entre les niveaux multistrukturés et relationnels de la compréhension, l'enseignant peut planifier des discussions en petits groupes où les élèves débattent des mérites respectifs de leurs idées et suggèrent des relations entre les concepts étudiés (Trudel, 2005). Dans une perspective à plus long terme, l'utilisation de modèles de réponses à l'item, tel

le modèle de l'échelle d'appréciation, peut permettre de construire un grand nombre de super-items SOLO partageant la même échelle de mesure et portant sur différents domaines de la physique au secondaire, ce qui faciliterait le choix de super-items adaptés au cheminement de l'élève (Ingebo, 1997).

NOTES

1. Le logit est l'unité probabiliste utilisée pour calculer la capacité des personnes et la difficulté des items (Penta, Arnould & Decruynaere, 2005, p. 37).
2. Nous n'avons pas inclus dans le test de questions sur le plan étendu-abstrait pour des raisons qui seront présentées plus loin.
3. Il ne semble pas que les différences dans les temps alloués à la passation du test aient influencé les résultats puisque les super-items de la première moitié du test (1 à 8) démontrent d'aussi bons degrés d'accord avec le modèle de l'échelle d'appréciation que les super-items de la seconde moitié (9 à 16) (voir tableau 2).
4. L'infit est un indice d'ajustement pondéré des items ou des personnes aux prescriptions du modèle choisi (Penta, Arnould & Decruynaere, 2005, pp. 94-95).
5. Le choix de l'origine de l'échelle de mesure est arbitraire. En effet, cette origine peut être centrée sur la moyenne des degrés de difficulté des super-items ou des degrés d'habiletés des élèves. Quel que soit le cas, c'est la différence entre le degré d'habileté d'un élève et le degré de difficulté du super-item qui nous permet de calculer la probabilité de réussite (Embretson & Reise, 2000).
6. L'étiquette «préstructuré» ne correspond pas à un niveau défini par la taxonomie SOLO. Cette étiquette permet seulement de classer les élèves qui n'ont réussi aucune question d'un super-item particulier (Biggs & Collis, 1982).

RÉFÉRENCES

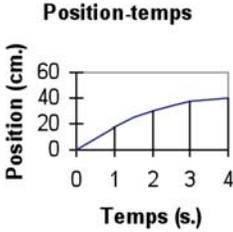
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park: Sage Publications.
- Arons, A.B. (1990). *A guide to introductory physics teaching*. Toronto: John Wiley & Sons.
- Baillargeon, G. (1984). *Techniques statistiques*. Trois-Rivières (Québec): Les Éditions SMG.
- Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Biggs, J. (2003). *Teaching for quality learning at university: What the student does* (2^e édition). Berkshire (United Kingdom): Open University Press.
- Biggs, J., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Toronto: Academic Press.
- Bliss, J., Morrison, I., & Ogborn, J. 1988. A longitudinal study of dynamics concepts. *International Journal of Science Education*, 10(1), 99-110.
- Bond, T.G., & Fox, C.M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Collis, K.F., & Biggs, J.B. (1991). Developmental determinants of qualitative aspects of schooling. In G. Evans (éd.), *Learning and teaching cognitive skills* (pp. 185-207). Victoria (Australie): Australian Council for Educational Research (ACER).
- Collis, K.E., & Davey, H.A. (1986). A technique for evaluating skills in high school science. *Journal of Research in Science Teaching*, 23(7), 651-663.

- Dickie, L.O. (1988). *Preconcepts in physics* (Rapport au John Abbott College Research and Development Committee). East Lansing, MI: National Center for Research on Teacher Learning (Base de données ERIC n° ED318533).
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologist*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Haladyna, T.M. (2004). *Developing and validating multiple-choice test items* (3^e édition). Hillsdale (NJ): Lawrence Erlbaum Associates.
- Halloun, I. (1996). Schematic modeling for meaningful learning of physics. *Journal of Research in Science Teaching*, 33(9), 1019-1041.
- Halloun, I.A., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53(11), 1043-1055.
- Hattie, J., & Purdie, N. (1998). The Solo model: Addressing fundamental measurement issues. In B. Hart & G. Boulton-Lewis (éds), *Teaching and learning in higher education* (pp. 145-176). Australian Council for Educational Research: Acer Press.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30, 141-158.
- Huffman, D., & Heller, P. (1995). What does the force concept inventory actually measure? *The Physics Teacher*, 33, 138-143.
- Ingebo, G.S. (1997). *Probability in the measure of achievement*. Chicago (IL): MESA Press.
- Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests en sciences humaines* (2^e édition). Paris: De Boeck Université.
- Legendre, M.-F. (2002). Le rôle du raisonnement qualitatif dans les processus de changement conceptuel et ses implications pour l'enseignement et la formation des enseignants. In R.M.J. Toussaint (éd.), *Changement conceptuel et apprentissage des sciences : Recherches et pratiques* (pp. 177-201). Outremont (Québec): Les Éditions LOGIQUES.
- Linacre, J.M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J.M., & Wright, B.D. (1993). *A user's guide to Facets: Rasch measurement computer program*. Chicago: Mesa Press.
- Masters, G.N., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In Norman Frederiksen, Robert J. Mislevy & Isaac I. Bejar (éds), *Test theory for a new generation of tests* (pp. 219-242). Hillsdale (NJ): Lawrence Erlbaum Associates.
- Oliva, J.M. (1999). Structural patterns in students' conceptions in mechanics. *International Journal of Science Education*, 21(9), 903-920.
- Penta, M., Arnould, C., & Decruynaere, C. (2005). *Développer et interpréter une échelle de mesure : Applications du modèle de Rasch*. Sprimont (Belgique): Pierre Mardaga éditeur.
- Sadler, P.M. (2000). The relevance of multiple-choice testing in assessing science understanding. In J.J. Mintzes, J.H. Wandersee & J.D. Novak (éds), *Assessing science understanding: A human constructivist view* (pp. 249-278). Toronto: Academic Press.
- Siegel, S., & Castellan, N.J., Jr. (1988). *Nonparametric statistics for the behavioral sciences* (2^e édition). Montréal: McGraw-Hill.
- Stouffer, S.A., Guttman, L., Suchman, E.A., Lazarfeld, P.F., Star, S.A., & Clauser, J.A. (1950). *Measurement and prediction*. New York: John Wiley & Sons.

- Treagust, D.F. (1995). Diagnostic assessment of students' science knowledge. In S.M. Glynn & R. Duit (éds), *Learning science in the schools: Research informing practice* (pp. 327-346). Mahwah (NJ): Lawrence Erlbaum Associates.
- Trowbridge, D.E., & McDermott, L. (1980). Investigation of student understanding of the concept of velocity in one dimension. *American Journal of Physics*, 48(12), 1020-1028.
- Trowbridge, D.E., & McDermott, L. (1981). Investigation of the concept of acceleration in one dimension. *American Journal of Physics*, 49(3), 242-253.
- Trudel, L. (2005). *Impact d'une méthode de discussion sur la compréhension des concepts de la cinématique chez les élèves de cinquième secondaire*. Thèse de doctorat, Université du Québec à Montréal, Montréal.
- White, R., & Gusntone, R. (1992). *Probing understanding*. London: The Falmer Press.
- Wilson, M. (1989). A comparison of deterministic and probabilistic approaches to measuring learning structures. *Australian Journal of Education*, 33(2), 127-140.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah (NJ): Lawrence Erlbaum Associates.
- Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis*. Chicago: Mesa Press.
- Wright, B.D., & Stone, M.H. (1979). *Best test design*. Chicago: Mesa Press.

ANNEXE

5. Le graphique suivant décrit la position d'une bille en cm à toutes les secondes lorsqu'elle remonte une légère pente. Après avoir monté sur une distance de 40 cm, la bille redescend la même pente.



5.1 Sur quelle distance la bille monte-t-elle avant de redescendre ?

Rép. : _____.

5.2 Quelle est son déplacement entre $t = 2$ s et $t = 4$ s ?

Rép. : _____.

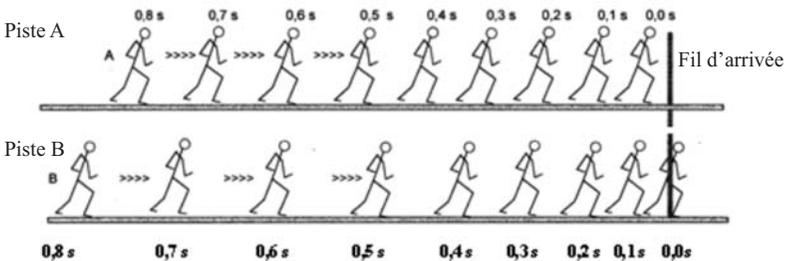
5.3 Quel est sa vitesse moyenne entre $t = 2$ s et $t = 4$ s ?

Rép. : _____.

.....

.....

12. Lors de la dernière seconde d'une course de cent mètres, une caméra ultra-rapide enregistre la position de deux coureurs, le coureur A et le coureur B, à tous les dixièmes de seconde. L'analyse des photographies montre que les deux coureurs ont ralenti peu avant de franchir le fil d'arrivée.



12.1 Quel coureur arrive le premier au fil d'arrivée?

Encerclez la bonne réponse.

- a) Le coureur A est arrivé le premier.
- b) Le coureur B est arrivé le premier.
- c) Le coureur A et le coureur B sont arrivés ensemble au fil d'arrivée.

12.2 À 0,7 seconde de la fin de la course, lequel des deux coureurs était en avant de l'autre? Encerclez la bonne réponse.

- a) Les deux coureurs étaient côte à côte.
- b) Le coureur A était en avant.
- c) Le coureur B était en avant.

12.3 Lequel des deux coureurs a ralenti le plus rapidement lors de cette fin de course? Encerclez la bonne réponse.

- a) Les deux coureurs ont ralenti aussi rapidement l'un que l'autre.
- b) Le coureur A a ralenti le plus rapidement.
- c) Le coureur B a ralenti le plus rapidement.