

La constance des erreurs dans une épreuve de rendement sur le concept

Jeannine Lavoie-Sirois
université laval

Dans cette recherche, l'auteure a vérifié la consistance des réponses accordées aux items semblables d'une épreuve sur l'identification des exemples d'un concept. Un concept géométrique et un concept grammatical ont fait l'objet de deux études indépendantes auprès d'élèves de niveau Secondaire III. L'indice de consistance individuelle (ICI) de Tatsuoka et Tatsuoka (1980) a été appliqué pour vérifier l'hypothèse que les erreurs commises étaient constantes. Les résultats obtenus ont confirmé cette hypothèse dans l'un et l'autre cas.

In this research, the writer has verified the consistency of responses to similar items of a test on identifying the examples of a concept. A geometric and a grammatical concept were the subjects of two independent studies on students in third year high school. The individual consistency index (ICI) of Tatsuoka and Tatsuoka (1980) was used to prove the hypothesis that the errors made were constant. The results obtained confirmed the hypothesis in both cases.

En vérifiant certaines hypothèses du modèle proposé par Markle et Tiemann (1969) pour guider l'enseignement et la mesure d'un concept, quelques chercheurs ont effectué des recherches intéressantes sur les types d'erreurs que des élèves commettent dans une tâche d'identification. Dans leurs travaux, Swanson (1972), Feldman (1972), Tennyson, Woolley et Merrill (1972) et Girard (1979) ont ainsi démontré que la surgénéralisation ou la sous-généralisation d'un concept sont reliées au type ou au nombre d'exemples présentés lors de l'enseignement. Dans ces études, l'accent a été mis sur l'aspect quantitatif des erreurs observées. Tennyson et ses collaborateurs ont toutefois débordé le cadre des considérations quantitatives en retenant comme variables quelques caractéristiques intrinsèques des exemples et des non-exemples présentés dans la séance d'apprentissage.

Les écrits ne rapportent cependant pas de travaux axés expressément sur les composantes qualitatives des erreurs relevées dans une tâche d'identification. Cet état de fait explique la persistance de certaines interrogations face à la non-maîtrise d'un concept. Ainsi, peut-on affirmer qu'il existe un lien entre les différentes erreurs qu'un élève commet relativement à un concept déterminé? Peut-on également affirmer que ces erreurs sont reliées à une perception inexacte du rôle

joué par l'une ou l'autre des propriétés de ce concept? En rapport avec la répétition des erreurs, peut-on croire qu'un élève qui identifie mal un exemple particulier d'un concept tend à se méprendre chaque fois qu'un exemple semblable lui est présenté au cours d'un même exercice?

Le but de la présente étude a été d'apporter un éclairage relatif à la répétition des erreurs caractéristiques dans une tâche d'identification. Plus explicitement, on a cherché à savoir si les erreurs commises relativement à un concept sont des erreurs constantes. Sur le plan pédagogique, une réponse positive à cette question peut avoir un impact important. En effet, si on démontre que les élèves qui ne maîtrisent pas un concept tendent à répéter leurs erreurs, on en déduit alors que ces erreurs ne sont pas le fruit du hasard et traduisent vraisemblablement une représentation inadéquate du concept qui tend à persister. Démontrer la constance des erreurs commises relativement à un concept, c'est donc faire ressortir implicitement la nécessité de remédier aux difficultés d'apprentissage qui peuvent les expliquer.

MÉTHODOLOGIE

L'hypothèse de la recherche était qu'à plusieurs répliques d'une tâche d'identification, l'élève qui ne maîtrise pas le concept mesuré tend à répéter les mêmes erreurs. Un concept géométrique (hauteur du triangle) et un concept grammatical (adjectif verbal épithète) ont été retenus pour vérifier cette hypothèse. Chaque concept a fait l'objet d'une étude indépendante. L'expérimentation a été menée auprès de 595 élèves de niveau Secondaire III de la région de Québec.

Une approche systématique a été utilisée pour développer les items de la tâche d'identification qui serait soumise aux sujets de l'expérimentation. Des exemples et des non-exemples caractéristiques ont ainsi pu être construits pour représenter chacun des concepts étudiés. Pour être en mesure d'identifier des erreurs particulières, il fallait en effet distinguer les composantes intrinsèques de chacun des items de l'épreuve. C'est la méthode analytique de Markle et Tiemann (1969) qui a été appliquée pour définir les concepts. Cette méthode a permis de mettre en évidence, pour chacun des concepts, deux caractéristiques essentielles (*les propriétés critiques*) et deux attributs variables (*les propriétés non critiques*). Ainsi, dans le cas de la hauteur du triangle, la définition du concept se présentait comme suit:

1. *La hauteur du triangle* est une droite
 - a) issue d'un des sommets du triangle (*propriété critique*)
 - b) et abaissée perpendiculairement sur le tracé du côté opposé à ce sommet (*propriété critique*).

2. L'*emplacement* de cette droite peut varier (*propriété non critique*):
 - a) droite intérieure
 - b) droite extérieure.
3. L'*orientation* de cette droite peut également varier (*propriété non critique*):
 - a) droite verticale
 - b) droite horizontale.

Dans le cadre de définition adopté, chaque propriété présentait deux possibilités mutuellement exclusives. En effet, dans un exemple déterminé, une propriété critique était soit présente, soit absente, tandis qu'une propriété non critique ne pouvait présenter qu'un des deux éléments de variation retenus. L'approche de Guttman et Schlesinger (1966) a servi de modèle pour combiner entre eux les éléments des quatre propriétés impliquées dans la définition. Le tableau 1 donne les 16 structures fondamentales (A à P) qui sont ressorties des combinaisons effectuées.

Chacun des exemples de la tâche d'identification a été construit pour illustrer une des 16 structures présentées dans le tableau 1. Ainsi, un exemple développé sur la base de la structure C (1, 2, 3b, 4a) pouvait donner la figure 1.

Afin de disposer de données qui permettraient de vérifier si une erreur était répétée, trois exemples ont été développés pour illustrer chacune des 16 structures fondamentales. Les différences apportées aux exemples représentatifs d'une même structure étaient mineures. Ainsi, dans le cas de la hauteur du triangle, ces différences portaient sur la grandeur de la figure, sur l'inversion de l'image et sur la dénomination des droites. Les trois séries de ces exemples ont constitué un trio de sous-tests semblables administrés l'un à la suite de l'autre au cours d'une seule séance d'expérimentation. A chacun des items de la tâche, les sujets devaient indiquer si l'exemple soumis représentait ou non le concept mesuré. Un ordre au hasard différent distribuait les items semblables dans chacun des sous-tests.

L'élaboration systématique des items de l'épreuve visait strictement à permettre l'analyse et la comparaison de caractéristiques déterminées des erreurs observées. Pour les fins de l'étude, le nombre de propriétés observées ou le nombre d'éléments retenus pour les attributs variables importaient donc peu. Une attention particulière a par ailleurs été portée au choix tant de ces attributs variables que de leurs éléments. Des enseignant(e)s en mathématique ou en français du cours secondaire ont contribué à identifier les propriétés non critiques qui semblaient pertinentes à retenir dans la définition des concepts étudiés.

La vérification de l'hypothèse a été effectuée en analysant la consistance

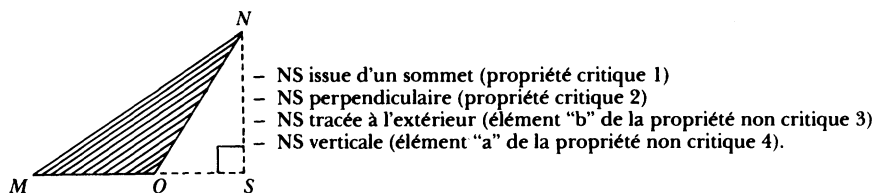


FIGURE 1
Un exemple de la tâche d'identification

TABLEAU 1
Structures obtenues en combinant les éléments
des propriétés

Structures fondamentales	Caractéristiques ^a			
	Propriétés critiques		Propriétés non critiques	
A	1	2	3a	4a
B	1	2	3a	4b
C	1	2	3b	4a
D	1	2	3b	4b
E	1	-2	3a	4a
F	1	-2	3a	4b
G	1	-2	3b	4a
H	1	-2	3b	4b
I	-1	2	3a	4a
J	-1	2	3a	4b
K	-1	2	3b	4a
L	-1	2	3b	4b
M	-1	-2	3a	4a
N	-1	-2	3a	4b
O	-1	-2	3b	4a
P	-1	-2	3b	4b

^aLégende:

1 (ou 2): présence de la propriété critique 1 (ou 2)

-1 (ou -2): absence de la propriété critique 1 (ou 2)

3a (ou 4a): la propriété non critique 3 (ou 4) présente l'élément "a"

3b (ou 4b): la propriété non critique 3 (ou 4) présente l'élément "b"

des patrons de réponses de chaque sujet aux trois sous-tests de l'épreuve. Le patron de réponses se définit ici comme un vecteur ou une suite ordonnée de succès/échecs notés selon le code binaire 1 ou 0. La stabilité des réponses apportées aux items semblables de l'épreuve a été déterminée par la valeur de l'indice de consistance individuelle ICI de Tatsuoka et Tatsuoka (1980). Pour vérifier l'hypothèse avancée, une étude a été faite de l'ensemble des valeurs de cet indice pour les sujets qui avaient commis au moins une erreur dans la tâche d'identification. Le degré de consistance des patrons de réponses d'un élève aux trois sous-tests de l'épreuve impliquait nécessairement le degré de consistance des réponses erronées que comportaient ces patrons. Dans les procédures de vérification de l'hypothèse, la distinction n'a toutefois pas été faite entre la consistance des erreurs d'exclusion (exemples rejetés à tort de la classe du concept) et la consistance des erreurs d'inclusion (non-exemples acceptés à tort dans la classe du concept).

Avant de présenter les résultats des deux études, expliquons brièvement l'indice de consistance individuelle ICI de Tatsuoka et Tatsuoka (1980) et illustrons la procédure de calcul qui a été appliquée pour déterminer cet indice chez les sujets observés.

Le propre de l'indice ICI est de déterminer la consistance des patrons de réponses fournis par un même élève à des épreuves répétées. La valeur d'un ICI permet donc de préciser jusqu'à quel point un élève tend à répondre de la même façon aux items d'une épreuve, c'est-à-dire tend à échouer ou à réussir les mêmes items d'une occasion à une autre. L'application de la formule du ICI requiert au départ que l'élève présente au moins deux réponses à chacun des items d'un test de rendement. De plus, les résultats obtenus à ces items doivent être codés selon le mode binaire 1 ou 0. Les valeurs possibles d'un ICI vont de -1 à $+1$. Plus un élève répond de façon consistante aux items qui sont semblables, plus il obtient un ICI qui se rapproche de 1.

Pour comparer les différents patrons de réponses de chaque sujet interrogé dans l'une ou l'autre étude, les résultats (1, 0) aux trois séries d'items ont été disposés selon l'ordre décroissant de difficulté de ces items pour le sujet. Le tableau 2 donne une illustration de la procédure qui a été appliquée pour obtenir le ICI de chaque sujet de l'expérimentation suite à ses réponses aux trois sous-tests semblables de l'épreuve.

Dans les cas où tous les patrons de réponses comparés ne présentent que des succès ou que des échecs (donc uniquement des 1 ou des 0), la consistance démontrée est évidemment parfaite et le ICI doit prendre logiquement la valeur 1. Cependant, l'application de l'expression $2Ua/U$ conduit à une division par zéro lorsque U est nul (ce qui se produit quand $n = x$ ou quand $x = 0$ puisque $U = x(n - x)$). Dans ce cas particulier,

TABLEAU 2

*Illustration de la démarche suivie pour le calcul du ICI d'un sujet***1 Résultats du sujet aux 16 items des trois sous-tests**

Items:	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Sous-test 1:	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1
Sous-test 2:	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1
Sous-test 3:	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1

2 Présentation des résultats selon l'ordre décroissant de difficulté des items pour le sujet

Items:	B	K	E	G	L	A	C	D	F	H	I	J	M	N	O	P
Sous-test 1:	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1
Sous-test 2:	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Sous-test 3:	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1

3 Coordonnées pour le calcul du ICI

S = un vecteur patron de réponses (1, 0) présentant en rangée et de gauche à droite les succès/échecs du sujet aux 16 items A à P selon leur ordre décroissant de difficulté pour ce sujet

n = le nombre d'items d'un vecteur S

x = le nombre d'items réussis dans un vecteur S

$U = x(n - x)$ d'un vecteur S

Ua = la somme des résultats obtenus en multipliant chaque élément "1" d'un vecteur S par le nombre de "0" le précédant de gauche à droite

Ainsi, pour le sous-test 1 de cet exemple,

$$Ua = (1 \times 2) + 11(1 \times 4) = 46$$

W = le rapport entre le U d'un vecteur S et la somme des U de tous les vecteurs impliqués

Dans ce cas-ci, pour le vecteur S -1,

$$W = \frac{U1}{U1 + U2 + U3} = \frac{48}{48 + 39 + 39} = .381$$

$$C_p(O) = \text{la consistance des patrons (vecteurs) selon un ordre déterminé}$$

$$= \frac{2Ua}{U} - 1$$

TABLE 1 (Concluded)

4 Résultats des calculs pour déterminer l'indice ICI du sujet

Vecteurs	$n = 16$	x	U	W	Ua	$C_p(O)$	$W.C_p(O)$
S-1	0010011111111111	12	48	.381	46	0.917	0.349
S-2	0001111111111111	13	39	.310	39	1.000	0.310
S-3	0001111111111111	13	39	.310	39	1.000	0.310
TOTAL	0012233333333333	38	126				0.969

L'indice ICI de ce sujet vaut donc 0.969.

l'application de la formule du ICI ne peut donc indiquer la constance parfaite qu'on observe pourtant dans les réponses de l'élève. Pour remédier à cet inconvénient, les auteurs proposent ou bien de décréter à priori que le ICI de ces élèves vaut 1, ou bien de modifier légèrement l'expression $2Ua/U$ pour empêcher l'annulation du dénominateur U . Une telle modification transforme alors la formule de $C_p(O)$ qui devient:

$$C_p(O) = \frac{(2Ua + 1)}{(U + 1)} - 1$$

RÉSULTATS ET ANALYSE

L'indice ICI a été calculé pour 228 sujets dans l'étude sur la hauteur du triangle et pour 285 sujets dans l'étude sur l'adjectif verbal épithète. Les figures 2 et 3 ci-après donnent la distribution des valeurs regroupées de l'indice ICI dans chacune des études. Il est intéressant de constater que, dans les deux cas, la distribution des scores est négativement dissymétrique. On observe donc que les indices élevés dominent, qu'il s'agisse du concept géométrique ou du concept grammatical.

Dans la figure 2, on indique un ICI moyen de 0,94 pour les sujets de l'étude sur la hauteur du triangle. Ce résultat apparaît élevé si on le compare au seuil minimum de 0,90 retenu par Tatsuoaka (1983) pour déclarer que la valeur du ICI est appréciable. Si on tient compte en plus de la valeur de la médiane ($Md = 0,97$) et du fait que près de 4 sujets sur 5 (79,4%) ont obtenu un ICI égal ou supérieur à 0,90, il semble bien qu'une proportion importante des sujets interrogés sur le concept géométrique ont présenté des patrons de réponses consistants aux trois sous-tests de l'épreuve. Par conséquent, les erreurs d'exclusion ou d'inclusion impliquées dans ces patrons de réponses ont été relativement stables. Il

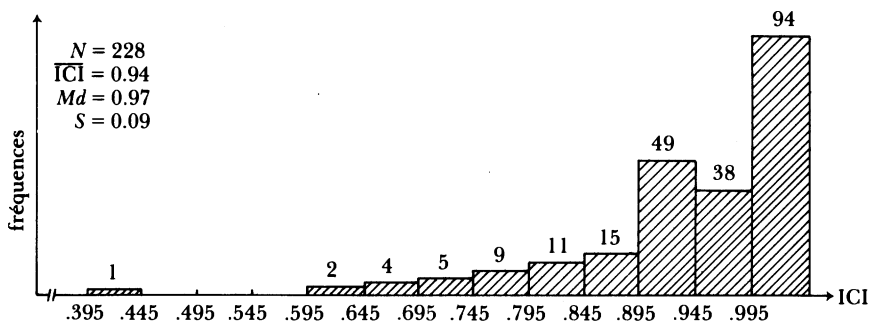


FIGURE 2

Histogramme des fréquences des valeurs ICI pour les sujets qui ont commis des erreurs dans la tâche d'identification sur la hauteur du triangle

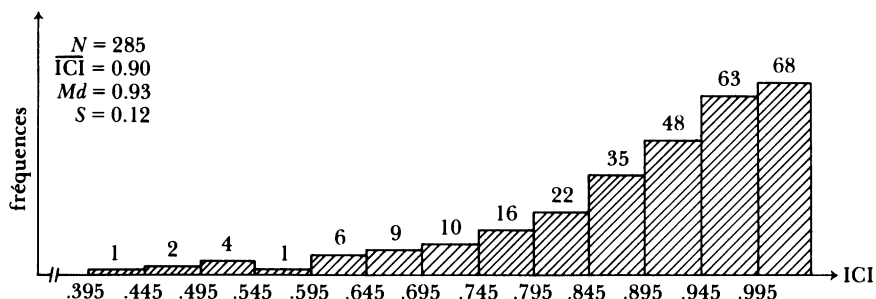


FIGURE 3

Histogramme des fréquences des valeurs ICI pour les sujets qui ont commis des erreurs dans la tâche d'identification sur l'adjectif verbal épithète

apparaît donc que l'hypothèse de la constance des erreurs tend à être confirmée dans l'étude sur la hauteur du triangle.

Dans l'étude sur l'adjectif verbal épithète, les sujets mesurés obtiennent un ICI moyen de 0,90. Toujours en se référant au seuil fixé par les auteurs pour déterminer un ICI acceptable, on constate à nouveau l'obtention d'une valeur moyenne intéressante. On calcule par ailleurs une proportion de 62,8% des sujets au-dessus du seuil minimum de 0,90. Cette proportion est donc moindre que dans la première étude. Si on considère toutefois que le ICI moyen du groupe demeure élevé, que la valeur de la médiane est importante ($Md = 0,93$) et qu'un plus grand nombre de ICI se retrouvent au-dessus qu'au-dessous du seuil de 0,90, il apparaît cette fois encore que les patrons de réponses aux trois sous-tests de l'épreuve

tendent à être consistants. Les erreurs d'exclusion ou d'inclusion présentes dans ces patrons paraissent donc constantes, d'où la confirmation à nouveau de l'hypothèse vérifiée.

Dans chacune des deux études, à partir de critères bien définis, on a donc démontré que les sujets observés tendaient à répéter leurs erreurs caractéristiques d'exclusion ou d'inclusion. On retiendra cependant que cette tendance a été plus marquée dans l'étude sur le concept de hauteur du triangle. Comme les sujets interrogés sur ce dernier concept ont davantage réussi la tâche d'identification que ceux interrogés sur l'adjectif verbal épithète (des résultats respectifs moyens de 85,04% et 58,13%), on peut se demander si la consistance des réponses aux items d'épreuves semblables n'est pas reliée au score total à l'ensemble de ces épreuves. On sait que Tatsuoka (1983) s'est intéressé à la relation qui peut exister entre le ICI et le score total d'un élève. En analysant les réponses fournies à des items semblables sur l'addition ou la soustraction d'entiers relatifs, il a observé que les erreurs constantes étaient surtout commises par les élèves qui obtenaient à la fois un ICI élevé (supérieur à 0,90) et un score total bas (inférieur à 80%). Par ailleurs, les erreurs non constantes étaient relativement plus nombreuses chez les élèves dont le ICI et le score total étaient tous deux peu élevés. Il serait intéressant que d'autres études explorent cette relation qui semble exister entre le résultat global à des mesures répétées et la stabilité des erreurs commises.

Notons que le petit nombre de concepts étudiés, l'administration de seulement trois sous-tests semblables et l'observation de sujets représentant un seul niveau d'études sont des facteurs qui limitent au départ la portée des résultats de cette recherche. Relativement à la constance des erreurs, on ne saurait donc généraliser les conclusions tirées de ces résultats pour tous les concepts ou pour tous les élèves.

Précisons en plus que le seuil à partir duquel une valeur ICI est estimée "élevée" demeure un facteur qui peut modifier plus ou moins les résultats d'une recherche comme celle-ci. Ainsi, si on avait exigé un seuil minimum de 0,88 ou de 0,92 pour admettre la consistance des patrons de réponses présentés par les sujets de l'expérimentation, la proportion de patrons jugés stables aurait pu varier sensiblement. Même si Tatsuoka (1983) a retenu un seuil minimum de 0,90, il semble bien que cette valeur a résulté d'un choix arbitraire de sa part. Dans l'appréciation des résultats de cette étude, il faut donc tenir compte de la délimitation arbitraire du seuil à partir duquel un indice ICI a été considéré comme significativement élevé.

Soulignons enfin deux variables non contrôlées qui peuvent avoir conduit à une sous-estimation de la tendance des sujets à répondre de la même façon à des items semblables. D'abord, chez certains élèves, il n'est pas impensable que l'observation répétée d'exemples ou de non-exemples

présentant les mêmes caractéristiques ait engendré une modification de l'apprentissage au cours de l'exercice de mesure. Des erreurs commises dans le premier sous-test peuvent alors avoir été évitées dans l'un ou l'autre des sous-tests subséquents. Si l'état de l'apprentissage différait entre le premier et le dernier sous-test, la valeur calculée de l'indice ICI ne pouvait que sous-estimer la tendance réelle à accorder une même réponse à des items semblables. Ce cas illustre la possibilité qu'un gain d'apprentissage au cours d'une épreuve peut contribuer à l'affaiblissement d'un indice de consistance. Il est également possible que, chez certains des sujets interrogés, une diminution de l'attention ou de l'intérêt se soit produite au fur et à mesure que les mêmes items caractéristiques étaient rencontrés. Dans un tel cas, le degré de consistance indiqué par le ICI peut avoir été inférieur à celui qu'on aurait observé si la concentration ou la motivation de ces élèves s'étaient maintenues du début à la fin de l'épreuve. On constate donc que des facteurs difficilement contrôlables peuvent contribuer à diminuer la valeur du ICI pour certains sujets. Il faut donc apprécier les résultats de l'étude en tenant compte de ces facteurs.

Face à la constance des erreurs commises dans l'identification des exemples d'un concept, d'autres variables pourraient être étudiées. Par exemple, on pourrait vérifier si l'âge de l'élève, la nature du concept mesuré, le nombre de patrons de réponses comparés ou la longueur de la tâche d'identification influencent la valeur d'un indice de consistance individuelle. On pourrait également vérifier si les réponses incertaines qui sont parfois accordées dans une tâche d'identification tendent à être constantes. En rapport avec l'interprétation de l'indice ICI, il serait de plus pertinent que l'on détermine scientifiquement le seuil à partir duquel la valeur d'un tel indice est significativement élevée. L'exploration du champ des erreurs dans une tâche d'identification ou des techniques appliquées pour étudier leurs caractéristiques n'en est qu'à ses débuts. Les résultats d'autres travaux ne peuvent donc qu'être souhaités si on veut mieux s'expliquer le phénomène de la non-maîtrise des concepts, et partant, être en mesure de remédier avec plus d'efficacité aux difficultés qui semblent à l'origine des erreurs commises.

RÉFÉRENCES

- Feldman, K. V. (1972). *The effects of number of positive and negative instances, concept definition and emphases of relevant attributes on the attainment of mathematical concepts* (Technical report). Madison, WI: Research and Development Center for Cognitive Learning.
- Girard, R. (1979). *La généralisabilité d'erreurs conceptuelles dans le cadre de l'évaluation formative de l'apprentissage de concepts*. Québec: Université Laval.

- Guttman, L., & Schlesinger, I. M. (1967). Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement*, 27, 569-588.
- Lavoie-Sirois, J. (1984). *La représentation des propriétés d'un concept et les problèmes de performance dans une tâche d'identification*. Québec: Université Laval.
- Markle, S. M., & Tiemann, P. W. (1969). *Really understanding concepts*. Champaign, IL: Stipes Publishing Company.
- Swanson, J. E. (1972). *The effects of number of positive and negative instances, concept definition and emphases of relevant attributes on the attainment of three environmental concepts by sixth-grade children* (Technical report). Madison, WI: Research and Development Center for Cognitive Learning.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.
- Tatsuoka, K. K., & Tatsuoka, M. M. (1980). *Detection of aberrant response patterns and their effect on dimensionality*. Urbana: University of Illinois, Computer-based Education Research Laboratory.
- Tennyson, R. D., Woolley, F. R., & Merrill, M. D. (1972). Exemplar and non-exemplar variables which produce correct concept classification behavior and specific classification errors. *Journal of Educational Psychology*, 63, 144-152.

Nach einer Bestimmung der Begriffsbeispiele, bestätigt der Autor in dieser Untersuchung die Kongruenz der Reaktionen auf ähnliche Probepunkte. Zwei unabhängige Untersuchungen prüften Studenten in ihrem dritten Hochschuljahr durch einen geometrischen und einen grammatischen Begriff. Zur Bestätigung der Hypothese, daß die gemachten Fehler konstant sind, wurde der ICI (individual consistency index = Individuelle Kongruenzindex) von Tatsuoka und Tatsuoka (1980) angewendet. In beiden Fällen wurde die Hypothese durch die Testergebnisse bestätigt.

En esta investigación, la autora verificó la consistencia de las respuestas dadas a ítems semejantes de una prueba sobre la identificación de los ejemplos de un concepto. Un concepto geométrico y un concepto gramatical fueron objeto de dos estudios independientes usando alumnos de 3er año de secundaria. Se aplicó el índice de consistencia individual (ICI) de Tatsuoka y Tatsuoka (1980) para verificar la hipótesis de que los errores cometidos eran constantes. Los resultados obtenidos confirmaron esta hipótesis en ambos casos.

Jeannine Lavoie-Sirois est professeure à la Faculté des sciences de l'éducation, Université Laval, Ste-Foy, Québec G1K 7P4.