

Analyse d'Erreurs Conceptuelles dans le Cadre de l'Evaluation Formative de l'Apprentissage de Concepts

Richard Girard, Marc-André Nadeau, et Gérard Scallon
université laval

Le but de cette étude est de démontrer que, dans un contexte d'évaluation formative, un instrument de mesure construit selon le modèle diagnostique développé par Suzan Markle et Philip Tiemann (1969) est suffisamment sensible et suffisamment fiable pour détecter des erreurs conceptuelles susceptibles d'être commises par un élève au début de l'apprentissage d'un concept. Les résultats obtenus auprès d'un échantillon de 180 élèves de sixième année, soumis à trois modalités d'enseignement de trois concepts différents, permettent de démontrer que les erreurs conceptuelles telles que la surgénéralisation et la sousgénéralisation sont étroitement associées à des caractéristiques spécifiques de l'enseignement. De plus, le lien entre des erreurs conceptuelles d'une part et des caractéristiques de l'enseignement d'autre part est relativement indépendant de facteurs tels que le type de concept enseigné ou le niveau d'habileté analytique des sujets.

The purpose of this study is to demonstrate, in the context of formative evaluation, that a measuring instrument constructed according to the diagnostic model developed by Suzan Markle and Philip Tiemann (1969) is sufficiently sensitive and sufficiently reliable to detect the conceptual errors likely to be made by a student at the beginning of concept learning. The results obtained from a sample of 180 Grade six students who were taught three different concepts under three modes of instruction, enabled us to demonstrate that conceptual errors, such as over-generalization or under-generalization are closely associated with the specific characteristics of the instruction. Moreover, the link between the conceptual errors on the one hand and the characteristics of instruction on the other hand is relatively independent of such factors as the type of concept taught or the level of analytic ability of the subjects.

LE CONTEXTE DU PROBLÈME

Depuis quelques années, l'évaluation du rendement scolaire ne se limite plus à son rôle de sanction à la fin d'une étape ou d'un cours, mais elle doit aussi intervenir à divers moments au cours de l'apprentissage de l'étudiant pour le guider dans son cheminement. En effet, il est important que le professeur et l'élève fassent le point à des moments déterminés pour décider si ce dernier peut poursuivre des apprentissages ultérieurs, ou s'il doit revenir sur certaines notions avant de passer à des apprentissages nouveaux. Cette décision de faire cheminer ou non l'étudiant rejoint ce que Bloom, Hasting et Madaus (1971) ont appelé l'évaluation formative de l'apprentissage.

L'évaluation formative est un type d'évaluation caractérisé par un ensemble de décisions pouvant modifier la progression de chaque élève

dans une démarche d'apprentissage. Lorsqu'un objectif pédagogique est maîtrisé par l'élève, celui-ci peut progresser vers d'autres objectifs à atteindre. Cependant, dans le cas d'une non maîtrise de l'objectif, la progression de l'élève est momentanément interrompue. C'est alors qu'il devient important, en évaluation formative, d'identifier les difficultés rencontrées par cet élève et d'en connaître les causes afin de prescrire un enseignement correctif approprié (Scallon, 1980). Or les instruments de mesure utilisés le plus souvent en évaluation formative fournissent suffisamment d'informations à la prise de décision mais ne sont guère diagnostiques quant à la nature même des difficultés rencontrées. Chester Harris (1974) est d'avis que pour remplir cette fonction diagnostique, un instrument de mesure doit être construit selon un modèle qui soit lui-même diagnostique.

Le modèle de Markle et Tiemann (1969) pour l'enseignement et la mesure des concepts fournit un cadre de référence pouvant intéresser le diagnostic pédagogique. L'originalité de ce modèle se situe au niveau de la mesure, et la tâche proposée à l'élève, pour vérifier son degré de maîtrise d'un concept, en est une de classification: l'élève doit identifier, parmi plusieurs exemples et non exemples d'un concept, ceux qui sont des exemples positifs de ce concept. L'hypothèse de Markle et Tiemann est qu'il est possible, au moment de corriger les réponses de l'élève, de dénombrer certaines erreurs de classification qui seraient associées à des caractéristiques particulières de l'enseignement.

Le but de la présente recherche est de valider le modèle de Markle et Tiemann c'est-à-dire de vérifier si les réponses données par des élèves dans une tâche de classification sont associées à certaines caractéristiques de l'enseignement.

DESCRIPTION DU MODÈLE DE MARKLE ET TIEMANN

Un concept est une classe d'objets ou d'événements qui possèdent une ou des caractéristiques communes que l'on appelle propriétés critiques ou essentielles. Ces objets constituent les exemples positifs du concept. Ces mêmes objets peuvent toutefois différer les uns des autres selon des caractéristiques non essentielles que l'on appelle propriétés non critiques. Un objet à qui il manque une ou plusieurs propriétés critiques est appelé exemple négatif du concept (Markle et Tiemann, 1969, p. 10).

Markle et Tiemann distinguent deux capacités requises à l'acquisition d'un concept. L'élève doit: (1) généraliser à des exemples positifs différents de ceux qui lui ont été présentés lors de l'enseignement; (2) discriminer les exemples positifs des exemples négatifs également différents de ceux présentés lors de l'enseignement.

Markle et Tiemann postulent que la capacité de l'élève à généraliser est liée à la présentation, lors de l'enseignement, d'exemples positifs suffisamment nombreux et variés pour mettre en évidence chacune des pro-

priétés non critiques (Feldman, p. 7). D'après les auteurs, le nombre d'exemples positifs dépend du nombre de propriétés non critiques susceptibles d'entraîner l'élève à commettre des erreurs. Tous les exemples positifs présentés lors de l'enseignement doivent provenir d'un ensemble planifié d'exemples (A rational set of examples, 1969, p. 21). Notons qu'il n'est pas exclu qu'une même propriété non critique soit mise en évidence plus d'une fois dans des exemples positifs différents.

Markle et Tiemann postulent également que la capacité de l'élève à discriminer est liée à la présentation, lors de l'enseignement, d'exemples négatifs suffisamment variés pour exclure chacune des propriétés critiques. Ici, le nombre d'exemples négatifs est déterminé par le nombre de propriétés critiques du concept étudié, et tous les exemples négatifs présentés lors de l'enseignement forment un ensemble planifié d'exemples négatifs (A rational set of non-examples, 1969, p. 21).

Le mode d'enseignement préconisé par Markle et Tiemann est plutôt de type inductif. L'élève reçoit les deux ensembles planifiés d'exemples positifs et d'exemples négatifs et, il doit découvrir par lui-même, pourquoi certains exemples présentés sont des exemples positifs du concept enseigné alors que d'autres ne peuvent être considérés comme tels. Le modèle de Markle et Tiemann repose sur l'hypothèse que la découverte, à partir d'exemples positifs et d'exemples négatifs, du caractère accidentel des propriétés non critiques et du rôle essentiel des propriétés critiques, est une stratégie d'enseignement suffisamment efficace pour amener l'élève à discriminer et à généraliser.

Au moment de l'évaluation, on demande à l'élève de classifier les exemples positifs et les exemples négatifs du concept en vue de vérifier sa capacité à généraliser et à discriminer. Les exemples positifs et les exemples négatifs présentés dans l'épreuve de rendement doivent être différents de ceux utilisés lors de l'enseignement et assez nombreux pour mettre tour à tour en évidence les propriétés critiques et les propriétés non critiques. L'étudiant capable d'identifier uniquement l'ensemble des exemples positifs présentés, démontre sa capacité à généraliser, alors que l'étudiant capable de reconnaître les exemples négatifs démontre aussi sa capacité à discriminer. L'élève qui est incapable d'identifier correctement les exemples positifs et les exemples négatifs présentés manifeste alors des difficultés d'apprentissage caractérisées. Ce second aspect du modèle de Markle et Tiemann sera développé dans les pages suivantes.

Markle et Tiemann affirment que l'élève auquel on n'a pas présenté lors de l'enseignement un ensemble planifié minimum d'exemples positifs et d'exemples négatifs est susceptible de commettre des erreurs de classification lorsqu'un lui demande d'identifier des exemples positifs et des exemples négatifs différents de ceux présentés lors de l'enseignement. Un premier type d'erreur de classification consiste à considérer comme positif un exemple négatif; ce type d'erreur est qualifié par les auteurs d'erreur de surgénéralisation (Overgeneralization) et ils postulent que la

surgénéralisation se produit lorsque les exemples négatifs présentés lors de l'enseignement ne sont pas suffisamment variés pour permettre à l'élève d'exclure chacune des propriétés critiques. Présenter alors des exemples plus discriminants serait, à leur avis, un enseignement correctif adéquat (1969, p. 11). L'élève peut commettre aussi l'erreur inverse consistant à ne pas reconnaître comme tels tous les exemples positifs présentés au moment de l'évaluation; l'élève classifiant comme négatifs des exemples positifs commet une erreur de sousgénéralisation (*undergeneralization*). Les erreurs de sousgénéralisation s'expliqueraient, selon Markle et Tiemann, par la présentation, lors de l'enseignement, d'un trop petit nombre d'exemples positifs suffisamment variés pour permettre à l'élève de découvrir le caractère accidentel ou non pertinent des propriétés non critiques. Comme enseignement correctif, il faudrait alors présenter à l'élève des exemples qui soient plus différents les uns des autres que ceux utilisés originalement (1969, p. 11).

Markle et Tiemann identifient une troisième façon d'être dans l'erreur: la confusion (*misconception*). L'élève confus n'arrive pas à identifier correctement tous les exemples positifs et tous les exemples négatifs qui lui sont présentés. Il comment à la fois les deux types d'erreurs décrits précédemment.

L'hypothèse avancée par les auteurs, comme étant à la source de la confusion, est l'absence combinée lors de l'enseignement d'exemples positifs et d'exemples négatifs suffisamment variés. Comme enseignement correctif, ils recommandent celui prévu lorsque l'élève commet des erreurs de surgénéralisation et de sousgénéralisation.

HYPOTHÈSES

Le modèle présenté par Markle et Tiemann suggère que l'apprentissage d'un concept est étroitement relié à la qualité et à la quantité d'exemple positif et d'exemple négatif présentés lors de l'enseignement de ce concept. Cet avancé est supporté par les résultats des recherches de Tennyson (1971), Markle et Tiemann (1972), Swanson (1972) et Feldman (1972). De plus, dans la présentation de leur modèle, Markle et Tiemann émettent des hypothèses concernant des caractéristiques d'enseignement susceptibles d'expliquer les erreurs commises par un étudiant incapable d'identifier des exemples positifs et des exemples négatifs de concepts différents de ceux utilisés lors de l'enseignement. Cette dimension du modèle de Markle et Tiemann n'a été que partiellement étudiée par Tennyson, Feldman et Swanson. Comme nous l'avons déjà affirmé, la possibilité d'utiliser le modèle de Markle et Tiemann à des fins diagnostiques dépend en grande partie de la validité du modèle au niveau de cette dimension.

Les lacunes d'enseignement ne représentent pas les seuls facteurs pouvant expliquer les erreurs conceptuelles. Selon les propos de Katherine

Feldman (1972, p. 86), les concepts eux-mêmes et des caractéristiques personnelles des élèves pourraient aussi constituer, à des degrés divers, des sources d'explications plausibles des erreurs conceptuelles. Si tel était le cas, la fiabilité des diagnostics posés dans une épreuve construite selon le modèle de Markle et Tiemann serait alors limitée à des concepts particuliers et à des élèves possédant certaines caractéristiques personnelles.

La perspective d'utiliser le modèle de Markle et Tiemann à des fins diagnostiques et prescriptives nous amène à formuler les hypothèses suivantes:

- H_{1a}) La présentation lors de l'enseignement d'exemples positifs et d'exemples négatifs peu variés et peu nombreux devrait être associée à un plus grand nombre d'erreurs de surgénéralisation.
- H_{1b}) Le lien entre des modalités d'enseignement et la commission d'erreurs de surgénéralisation devrait être stable peu importe le concept enseigné et peu importe le niveau d'habileté analytique des sujets.
- H_{2a}) La présentation lors de l'enseignement d'exemples positifs et d'exemples négatifs peu variés et peu nombreux devrait être associée à un plus grand nombre d'erreurs de sousgénéralisation.
- H_{2b}) Le lien entre des modalités d'enseignement et la commission d'erreurs de sousgénéralisation devrait être stable peu importe le concept enseigné et peu importe le niveau d'habileté analytique des sujets.

La confusion n'a pas été isolée comme variable dépendante puisqu'elle présente une combinaison des erreurs de surgénéralisation et des erreurs de sousgénéralisation. Un élève commet des erreurs de surgénéralisation ou de sousgénéralisation mais non de confusion. La confusion représente plutôt une façon pour l'étudiant d'être dans l'erreur qui se caractérise par une tendance marquée à commettre des erreurs de surgénéralisation et des erreurs de sousgénéralisation dans une même tâche de classification. Nous examinerons tout de même la confusion mais à titre exploratoire seulement.

LA MÉTHODOLOGIE

L'échantillon

Notre échantillon a été puisé dans quatre écoles élémentaires, à la Commission scolaire Belle-Rivière, dans la région du Lac St-Jean. Pour que les groupes soient comparables quant au rendement académique et quant aux méthodes d'enseignement utilisées par les professeurs, nous avons retenu les services de huit classes de sixième année pour un total de 180 élèves.

Le choix des concepts

Pour les fins de l'expérimentation, notre choix s'est arrêté sur trois concepts appelés: Oralfa, Symétrie Mixte et Métaphore. Quatre facteurs ont influencé le choix de ces concepts: (1) qu'ils soient nouveaux pour l'élève; (2) qu'ils ne soient pas trop complexes pour des élèves de sixième année; (3) que l'apprentissage de l'un ou de l'autre de ces concepts soit indépendant de l'apprentissage des deux autres concepts; et (4) que le contenu sur lequel porte le concept soit différent d'un concept à l'autre (le concept "Oralfa" est à contenu symbolique: groupement de lettres, le concept "Symétrie mixte" à contenu figuratif: figures et le concept "Métaphore" à contenu sémantique).

Le matériel

Le matériel utilisé comprend d'abord l'épreuve des figures cachées (Hidden Figures Test) de French, Ekstrom et Price (1963) qui nous a permis de classer les étudiants selon la dimension analytique-globale. Puis, pour chacun des trois concepts, nous avons élaboré trois documents d'entraînement en introduisant une variation quant au nombre d'exemples positifs et d'exemples négatifs présentés à l'élève pour un apprentissage par induction du concept. Cette variation définit trois modalités d'enseignement qui se présentent comme suite: (1) présentation à l'élève de huit exemples positifs et d'un seul exemple négatif; (2) présentation d'un seul exemple positif et de huit exemples négatifs; et (3) présentation d'un seul exemple positif et d'un seul exemple négatif. Précisons ici que la variation en nombre d'exemples introduite dans cette recherche s'inspire de la méthodologie employée dans plusieurs recherches en apprentissage de concepts (Bourne Jr., 1966 et Clark, 1971) et ne constitue pas une situation idéale ou maximale. L'objet de l'étude était d'éprouver la sensibilité d'un instrument de mesure construit selon le modèle de Markle et Tiemann et ce, selon trois modalités d'enseignement et à un moment où les élèves en sont à un point de départ et non pas à un point d'arrivée.

Une épreuve de rendement a été rédigée pour chacun des trois concepts. Chaque épreuve est composée de dix exemples positifs et de dix exemples négatifs différents de ceux utilisés lors de l'enseignement.

Le traitement

Pour répondre aux interrogations soulevées dans le cadre de cette recherche, nous avons adopté deux approches méthodologiques: l'analyse de la variance et la théorie de la généralisabilité. L'analyse de la variance permet de découvrir s'il existe une différence significative entre le nombre moyen d'erreurs de surgénéralisation ou de sousgénéralisation commises par les trois groupes de sujets soumis à des modalités d'enseigne-

TABLEAU 1
Analyse de la variance des résultats d'erreurs de surgénéralisation

<i>Sources de variation</i>	<i>Somme des carrés</i>	<i>Degrés de libertés</i>	<i>Carrés moyens</i>	<i>Rapports "F"</i>
Modalités				
d'entraînement (A)	201.25	2	100.62	5.62* (1)
Concepts (B)	85.44	2	42.72	3.69*
Degrés d'habileté				
analytique (C)	940.97	59	15.95	1.38
A × B	61.71	4	15.43	1.61
A × C	1424.31	118	12.07	1.25
B × C	1367.45	118	11.59	
A × B × C, e	2265.40	236	9.60	

* $p \leq .05$
(1) Quasi F

ment différentes. La théorie de la généralisabilité est surtout utilisée ici en vue de vérifier la fiabilité de l'effet des modalités d'enseignement sur le nombre d'erreurs de surgénéralisation et de sousgénéralisation commises par les différents groupes de sujets. La théorie de la généralisabilité (generalizability theory) développée par Lee J. Cronbach, Harinder Nanda, Goldine Gleser et Nageswari Rajaratnam en 1972, est utilisée en vue de déterminer les effets simples et combinés des facteurs "modalités d'enseignement," "concepts" et "degrés d'habileté analytique."

Les résultats

La vérification des hypothèses énoncées dans cette recherche a été réalisée en répondant aux deux questions suivantes formulées dans le cadre de la théorie de la généralisabilité: (1) Existe-t-il lien de dépendance entre le nombre d'exemples positifs et le nombre d'exemples négatifs qui définissent des modalités d'enseignement et le façon de répondre des élèves à une tâche de classification? (2) Le degré d'association qui pourrait être établi entre ces modalités d'enseignement et la façon de répondre des élèves dépend-t-il du concept enseigné (oralfa, symétrie mixte ou métaphore) ou du degré d'habileté analytique de ces élèves?

Les analyses de variance rapportées aux Tableaux 1 et 2 apportent une réponse positive à la première question. Les données rapportées au Tableau 1 montrent un quasi F de 5.62 significatif à .05 lorsque les modalités d'enseignement sont considérées comme source principale de variation. Ce résultat vient donc supporter l'hypothèse avancée à savoir qu'il existe un lien de dépendance entre le nombre d'erreurs de surgénéralisation commises par les élèves et les modalités d'enseignement

TABLEAU 2
Analyse de la variance des résultats d'erreurs de sousgénéralisation

Sources de variation	Somme des carrés	Degrés de liberté	Carrés moyens	Rapports "F"
Modalités				
d'entraînement (A)	216.04	2	108.02	16.12* (1)
Concepts (B)	74.01	2	37.01	4.99*
Degrés d'habileté				
analytique (C)	432.98	59	7.34	0.99
A × B	22.74	4	5.69	0.59
A × C	1246.62	118	10.56	1.11
B × C	874.66	118	7.41	
A × B × C	2254.59	236	9.55	

* $p \leq .01$
(1) Quasi F

auxquelles ils ont été soumis. De plus, les résultats de l'analyse de variance révèlent que les concepts considérés comme source principale de variation ont eux aussi un effet significatif sur le nombre d'erreurs de surgénéralisation ($F = 3.69, p \leq .05$). Cette constatation laisse présager que la nature des concepts utilisés au moment de l'enseignement doit être prise en considération dans une recherche comme la nôtre.

Les résultats à la variable sousgénéralisation s'interprètent de la même façon que ceux obtenus à la variable "surgénéralisation." En effet, les données du Tableau 2 nous révèlent un "quasi F" de 16.12 lorsque les modalités d'enseignement sont considérées comme source principale de variation, valeur dont le niveau de probabilité est inférieur à .01. Lorsque les concepts sont considérés comme source principale de variation, nous obtenons un rapport F de 4.99 soit une valeur dont le niveau de probabilité est aussi inférieur à .01. Ces premiers résultats nous amènent donc à reconnaître l'existence d'une différence dans le nombre d'erreurs de sousgénéralisation commises par les trois groupes de sujets. Ces résultats supportent l'hypothèse voulant que le fait de présenter, lors de l'enseignement d'un concept, des exemples positifs ou des exemples négatifs peu nombreux, ou de combiner ces deux traitements, a un effet différentiel sur le nombre d'erreurs de sousgénéralisation susceptibles d'être commises par les étudiants soumis à ces trois modalités d'enseignement. Cependant, les concepts représentent aussi un facteur qui introduit une différence significative entre les moyennes d'erreurs de sousgénéralisation commises par les étudiants.

Ces premiers résultats nous amènent donc à reconnaître l'existence d'une différence entre les groupes de sujets selon les modalités d'enseignement auxquelles ils ont été soumis. Cependant, ces résultats ne

TABLEAU 3
*Nombre moyen d'erreurs de surgénéralisation pour chacun et pour l'ensemble
des concepts et des groupes expérimentaux*

<i>Concepts</i>	<i>Groupes expérimentaux</i>			<i>N. moyen d'erreurs par concept</i>
	PEP	PEN	PEPN	
Oralfa	3.83	5.28	4.45	4.52
Symétrie mixte	2.80	5.27	3.72	3.93
Métaphore	3.23	4.78	4.25	4.09
N. moyen d'erreurs par groupe expérimental	3.29	5.11	4.14	

nous indiquent pas le sens des différences observées. Selon l'hypothèse de Markle et Tiemann, un nombre insuffisant d'exemples négatifs présentés lors de l'enseignement d'un concept serait associé à un plus grand nombre d'erreurs de surgénéralisation commises par des élèves dans une tâche de classification. D'autre part, et d'après les auteurs, un nombre insuffisant d'exemples positifs serait associé à un plus grand nombre d'erreurs de sousgénéralisation. Les statistiques portant sur le nombre moyen d'erreurs de chaque type commises par les élèves selon la modalité d'enseignement appuient cette double hypothèse (cf. Tableaux 3 et 4).

Les données du Tableau 3 démontrent effectivement que le groupe de sujets auquel on a présenté peu d'exemples négatifs lors de l'enseignement (PEN) est celui en moyenne commet le plus d'erreur de surgénéralisation sont 5.11 par rapport à 3.29 et 4.14 respectivement pour les deux autres groupes. De plus, cette tendance persiste pour n'importe lequel des trois concepts. Les données du Tableau 3 indiquent donc une différence dans le sens d'un nombre moyen d'erreurs de surgénéralisation plus élevé pour le groupe de sujets auquel peu d'exemples négatifs ont été présentés lors de l'enseignement.

Les valeurs rapportées au Tableau 4 s'interprètent dans le même sens. Ils révèlent que le groupe de sujets auquel on a présenté, lors de l'enseignement des trois concepts impliqués dans cette étude, peu d'exemples positifs (PEP), commet en moyenne plus d'erreurs de sousgénéralisation que les deux autres groupes soumis à des modalités différentes. Cette tendance observée pour l'ensemble des trois concepts ne vaut plus pour tous les concepts lorsque ceux-ci sont considérés séparément. Dans le cas du concept "ORALFA," le groupe PEPN (peu d'exemples positifs et peu d'exemples négatifs) commet en moyenne légèrement plus d'erreurs de sous-généralisation que le groupe PEP. Dans le cas du concept "Symétrie mixte," la différence en nombre moyen d'erreurs de sousgénéralisation est nulle entre ces deux groupes.

De l'ensemble de ces résultats se dégage une réponse positive à la première question formulée: il existerait donc un lien de dépendance

TABLEAU 4
*Nombre moyen d'erreurs de sousgénéralisation pour l'ensemble et pour chacun
des concepts et des groupes expérimentaux*

<i>Concepts</i>	<i>Groupes expérimentaux</i>			<i>N. moyen d'erreurs par concept</i>
	PEP	PEN	PEPN	
Oralfa	4.52	3.83	4.57	4.31
Symétrie mixte	4.65	2.82	4.65	4.04
Métaphore	4.38	3.10	3.57	3.68
N. moyen d'erreurs par groupe expérimental	4.52	3.25	4.26	

entre certaines modalités d'enseignement et le nombre d'erreurs de surgénéralisation et de sousgénéralisation commises par les élèves.

En ce qui a trait à la second question soulevée, les résultats de cette étude apportent une réponse négative: il semblerait en effet que le lien entre des modalités d'enseignement et la façon de répondre des élèves soit indépendant du concept enseigné et du degré d'habileté analytique de ces élèves. Cela amène à conclure qu'une épreuve construite selon le modèle de Markle et Tiemann permettrait de diagnostiquer de façon fiable des erreurs de surgénéralisation et/ou des erreurs de sousgénéralisation commises par des élèves peu importe le concept enseigné et peu importe le degré d'habileté analytique de ces mêmes élèves. Ce degré de fiabilité se traduit par un coefficient de généralisabilité. Ce dernier est considéré comme étant acceptable (selon Cardinet, Tourneur et Allal, 1976), s'il atteint un seuil de 0.80. Un tel coefficient de généralisabilité serait suffisant pour conclure que la différenciation introduite par les modalités d'enseignement est stable peu importe le concept enseigné, peu importe le degré d'habileté des sujets.

Les coefficients de généralisabilité rapportés au Tableau 5 sont de 0.81 lorsque le nombre d'erreurs de surgénéralisation est considéré comme variable dépendante et 0.90 lorsque le nombre d'erreurs de sousgénéralisation est considéré comme variable dépendante. Il semblerait donc que le lien existant entre des modalités d'enseignement et le nombre d'erreurs de surgénéralisation d'une part et le nombre d'erreurs de sousgénéralisation d'autre part soit indépendant des concepts enseignés et du degré d'habileté analytique.

Implications des résultats

Tentons maintenant de décrire certaines implications concernant les résultats rapportés dans les pages précédentes.

1 / La principale implication se rapporte à la fonction diagnostique d'une épreuve de rendement construite dans un contexte d'évaluation forma-

TABLEAU 5
*Estimés des composantes de variance et des coefficients de généralisabilité
des variables “surgénéralisation” et “sousgénéralisation”²*

Source	Estimés des composantes de variance	
	Surgénéralisation ³	Sousgénéralisation ⁴
Modalités d'enseignement (A)	0.46	0.56
Concepts (B)	0.17	0.16
Degrés d'habileté analytique	0.48	0.00 ¹
A × B	0.10	0.00 ¹
A × C	2.47	0.34
B × C	3.86	2.47
A × B × C, e	9.60	9.55

¹Valeurs faiblement négatives ramenées à zéro

²Nous avons retenu le dispositif expérimental dans lequel les modalités d'enseignement ont été considérées comme facette de différenciation tandis que les concepts et le degré d'habileté analytique ont été considérés comme facettes de généralisation aléatoire.

³Estimé des coefficients de généralisabilité ρ^2 (AB) = 0.81

⁴Estimé des coefficients de généralisabilité ρ^2 (AB) = 0.90

tive. Nous insistions au début de cette recherche pour qu'une telle épreuve offre la possibilité de nous prononcer non seulement sur le niveau de maîtrise de l'élève mais aussi sur les difficultés rencontrées en cours de route par ce même élève. Ainsi, les réponses fournies par un élève soumis à une épreuve comportant un ensemble d'exemples positifs et un ensemble d'exemples négatifs d'un concept, différents de ceux utilisés lors de l'enseignement, devraient refléter non seulement le niveau de maîtrise de cet élève (qui se traduit par le nombre de bonnes réponses) mais également le type de difficultés rencontrées par ce même élève. Ne pas accepter comme tels les exemples positifs nouveaux (erreurs de sousgénéralisation) traduit une difficulté chez l'élève à généraliser. Un enseignement comportant des exemples positifs peu variés et en quantité insuffisante serait responsable de cette situation. Par contre, l'élève qui n'identifie pas comme tels les exemples négatifs qui lui sont présentés (erreurs de surgénéralisation) épreuve manifestement des difficultés à discriminer. Ces erreurs de surgénéralisation s'expliqueraient par un enseignement comportant des exemples négatifs peu variés et en quantité insuffisante.

2 / La seconde implication de cette étude concerne la prescription faisant suite à un diagnostic posé. Puisque les erreurs de surgénéralisation ou les erreurs de sousgénéralisation commises par un élève sont le fruit d'un enseignement comportant certaines caractéristiques quant au nombre d'exemples positifs et d'exemples négatifs présentés, il semble implicite, selon le cas, que la présentation d'exemples positifs et d'exemples négatifs plus variés et plus nombreux corrigerait vraisemblablement la situation.

De plus, ce correctif pourrait être individualisé selon le type d'erreur qui caractérise l'élève. Comme la valeur prescriptive du modèle de Markle et Tiemann n'a pas fait spécifiquement l'objet de cette recherche, elle représente donc une hypothèse à vérifier dans une recherche ultérieure.

3 / La validation du modèle de Markle et Tiemann ne s'est pas limitée à établir expérimentalement le lien entre des erreurs conceptuelles typiques et des modalités d'enseignement et ce, à un moment où les élèves en sont à un point de départ dans l'apprentissage d'un concept. Nous devons nous assurer que des facteurs autres que des modalités d'enseignement comme le type de concept enseigné ou le degré d'habileté analytique des sujets ne représenteraient pas également des explications plausibles des erreurs conceptuelles commises par les élèves. Tels que rapportés antérieurement, les résultats démontrent qu'effectivement le lien entre les erreurs conceptuelles et les modalités d'enseignement est relativement stable peu importe le concept enseigné ou le type de sujets d'après la dimension analytique-globale. Les implications de ces résultats se situent au niveau de l'utilisation possible d'une épreuve de rendement construite selon le modèle de Markle et Tiemann. Que cette épreuve soit administrée à tout groupe de sujets quel que soit son niveau d'habileté analytique ou qu'elle porte sur des concepts différents en contenu, les résultats sont suffisamment fiables pour espérer que les erreurs conceptuelles décelées soient le résultat d'un enseignement comportant des exemples positifs et des exemples négatifs peu variés et en quantité insuffisante.

4 / Dans cette recherche, des modalités d'enseignement ont été introduites en variant le nombre d'exemples positifs et le nombre d'exemples négatifs présentés au moment de l'apprentissage. Il va de soi qu'aucune de ces modalités ne constitue un traitement idéal puisque le but de cette recherche n'était pas d'explorer les conditions maximales d'apprentissage d'un concept. Il faut comprendre que les tâches de classification ont été administrées à des élèves non pas au terme d'un apprentissage mais à un moment qui se situe au point de départ de cet apprentissage. Dans ce contexte, il semble qu'un instrument de mesure construit selon le modèle de Markle et Tiemann soit sensible à des différences introduites en modalités d'enseignement lorsque des élèves abordent l'apprentissage d'un concept qui leur est nouveau. Il serait intéressant d'explorer, dans des études ultérieures, la sensibilité de ce type d'instrument lorsqu'il est administré à divers moments dans une démarche d'apprentissage complète qui vise la maîtrise d'un concept.

Toutefois, le fait de pouvoir identifier et de corriger les erreurs de surgénéralisation et de sousgénéralisation pourrait résoudre pratiquement le problème de la confusion mais, le lien qui semble exister entre la surgénéralisation et la sousgénéralisation d'une part, et la confusion d'autre part, n'est pas évident. En somme, la confusion telle que décrite comme une manière d'être dans l'erreur qui soit distincte de la surgénéralisation et de la sousgénéralisation demeure une hypothèse.

Pour terminer, il est important de souligner que les conclusions de cette recherche s'inscrivent dans un cadre d'apprentissage par induction. Ce mode d'apprentissage, appliqué dans toute sa rigueur, se limite à présenter à l'étudiant des exemples positifs et des exemples négatifs d'un concept sans l'aide d'aucun indice verbal ni d'aucune définition du concept. Puisque vraisemblablement, il y a peu de concepts qui s'acquièrent par induction dans la pratique scolaire, on pourrait être tenté de réduire la portée de notre étude. Il nous apparaît alors important de signaler que des difficultés conceptuelles peuvent surgir, même à la suite d'un enseignement faisant usage d'exemples positifs, d'exemples négatifs, d'indices verbaux et de définitions. On ne peut affirmer pour l'instant que l'usage d'indices verbaux ou de définitions d'un concept au moment de l'apprentissage supplée à l'absence d'exemples positifs et d'exemples négatifs.

RÉFÉRENCES

- Airasian, P. W. Formative evaluation instruments. *The Irish Journal of Education*, 1968, 2, 127-135.
- Airasian, P. W., & Madaus, G. F. Functional types of student evaluation. *Measurement and Evaluation in Guidance*, 1972, 4, 221-223.
- Baker, E. Beyond objectives: Domain referenced tests for evaluation and instructional improvement. *Educational Technology*, juin 1974, 10-16.
- Bloom, B. S., Hasting, P. T., & Madaus, G. F. *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill, 1971.
- Bourne, Jr., L. *Human conceptual behavior*. Boston: Allyn & Bacon, Inc., 1966.
- Cardinet, J., & Tourneur, Y. *La généralisabilité des mesures du système scolaire*. Exposé présenté au "Second International Symposium on Educational Testing." Montreux, 1975.
- Cardinet, J., Tourneur, Y., & Allal, L. The symmetry of generalizability theory: Application to educational measurement. *Journal of Educational Measurement*, 1976, 13, 119-135.
- Clark, Cecil. Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. *Journal of Educational Psychology*, 1971, 62, 253-278.
- Cronbach, L. J., Gleser, G. C., Rajaratnam, N., & Nanda, H. *The dependability of behavioral measurements*. New York: John Wiley, 1972.
- Feldman, K. V. The effects of number of positive and negative instances, concept definition, and emphasis of relevant attributes on the attainment of mathematical concepts. Madison, Wisconsin: Research and Development Center for Cognitive Learning. Technical Report, no. 243, 1972.
- Frederick, W. C. *Information processing and concept learning at grades 6, 8 and 10 as function of cognitive style*. Madison, Wisconsin: Research and Development Center for Cognitive Learning, Tech. rep. No 44, 1968.
- French, J. W., Ekstrom, R. B., & Price, L. A. *A manual for kit of reference tests for cognitive factors*. Princeton, New Jersey: Educational Testing Service, 1963.
- Gillmore, G. M. *An introduction to generalizability theory as a contributor to evaluation research*. ERIC: Ed 178 599, Washington, 1979.
- Guttman, L., & Schlesinger, I. Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement*, 1967, 27, 569-580.

- Harris, C. W. Some technical characteristics of mastery tests. In M. Alkin, G. Harris, & J. Popham (Eds.), *Problems in criterion-referenced measurement*. CSE Monographs Series in Evaluation, no. 3. Los Angeles: Center for the Study of Evaluation, 1974.
- Hively, W., Patterson, H. L., & Page, S. H. A universe-defined system of arithmetic achievement test. *Journal of Educational Measurement*, 1968, 5, 275-290.
- Kane, Michael, T., Gillmore, G. M., & Crooks, T. J. Student evaluation of teaching: The generalizability of class means. *Journal of Educational Measurement*, 1976, 13(3).
- Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont: Brooks-Cole Publishing Company, 1968.
- Klausmeier, H. J., Ghatala, E. S., & Frayer, D. A. *Conceptual learning and development: A cognitive view*. New York: Academic Press, 1974.
- Madaus, G. F., & Airasian, P. W. Placement, formative, diagnostic, and summative evaluation of classroom learning. Paper presented to the 1970 Annual Meeting of the American Educational Research Association, Minneapolis. ERIC no. ED 041829, 1972.
- Markle, S. M., & Tiemann, P. W. *Really understanding concepts or in fruitless pursuit of the jabberwock*. Champaign, Illinois: Stiper, 1969.
- Millman, J. Criterion-referenced measurement. In J. Popham (Ed.), *Evaluation in education*. Berkeley: McCutchan Publishing Corporation, 1974, 311-397.
- Nitko, A. J., & Tse-Chi, Hsu. Using domain referenced test for student. Placement, diagnosis and attainment in a system of adaptive, individualized instruction. *Educational Technology*, juin 1974.
- Nitko, A. J. Distinguishing the many varieties of criterion-referenced tests. *Review of Educational Research*, 1980, 50, 461-485.
- Popham, J. W. Selecting objectives and generating tests items for objectives-based tests. In M. Alkin, C. Harris, J. Popham (Eds.), *Problems in criterion-referenced measurement*. CSE Monographs Series in Evaluation, no. 3. Los Angeles: Center for the Study of Evaluation, 1974.
- Scallon, G. *L'évaluation formative: Un outil au service de l'enseignement et de l'apprentissage*. Document non publié, 1980.
- Scriven, M. The methodology of evaluation. In R. W. Tyler, R. M. Gagné & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. AERA Monographs Series on Curriculum Evaluation, no. 1. Chicago: Rand McNally, 1967.
- Swanson, J. E. The effect of number of positive and negative instances, concept definition and emphasis of relevant attributes on the attainment of three environmental concepts of sixth-grade children. Madison, Wisconsin Research and Development Center for Cognitive Learning. Technical Report, no. 244, 1972.
- Tennyson, R. D. Correct classification, overgeneralization, undergeneralization, and misconception as a function of probability, divergency, and matching of instances in concept acquisition. Provo, Utah: Brigham Young University, Division of Communication Services Working Paper, no. 19, 1971.
- Tennyson, R. D., Wooley, F. R., & Merrill, M. D. Exemplar and nonexemplar variables which produce correct classification behavior and specified classification errors. *Journal of Educational Psychology*, 1972, 63, 144-152.
- Tourneur, Y., Frydman, M. *Une application de la théorie de la généralisabilité à l'estimation d'erreurs*. Université de l'Etat de Mons, Service d'étude des méthodes et des moyens d'enseignement, 1977.