Mesure et évaluation en éducation

Le modèle de la généralisabilité : une théorie de la mesure en éducation

Gratien Bambanota Mokonzi

Volume 26, numéro 1-2, 2003

Généralisabilité

URI : https://id.erudit.org/iderudit/1088236ar
DOI : https://doi.org/10.7202/1088236ar

Aller au sommaire du numéro

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)
2368-2000 (numérique)

Découvrir la revue

Citer cet article

Bambanota Mokonzi, G. (2003). Le modèle de la généralisabilité : une théorie de la mesure en éducation. *Mesure et évaluation en éducation*, *26*(1-2), 5–18. https://doi.org/10.7202/1088236ar

Résumé de l'article

Cet article présente la notion de généralisabilité, son origine dans les travaux psychométriques de Cronbach, son extension par symétrie à d'autres facettes que les questions, le déroulement d'une étude-G, ses résultats directs et leur utilisation possible pour améliorer la fidélité des mesures, dans la perspective des publications initiales de Cardinet, Tourneur et Allal.

Le modèle de la généralisabilité : une théorie de la mesure en éducation

Gratien Bambanota Mokonzi
Faculté de psychologie et des sciences de l'éducation, Université de Kisangani

MOTS CLÉS : Généralisabilité, ANOVA, modèle, mesure, dispositif, fidélité, variance d'erreur, optimisation

Cet article présente la notion de généralisabilité, son origine dans les travaux psychométriques de Cronbach, son extension par symétrie à d'autres facettes que les questions, le déroulement d'une étude-G, ses résultats directs et leur utilisation possible pour améliorer la fidélité des mesures, dans la perspective des publications initiales de Cardinet, Tourneur et Allal.

KEY WORDS : Generalizability, ANOVA, model, measurement, design, reliability, error-variance, D-study

This paper presents the concept of generalizability, its origin in Cronbach's psychometric papers, its extension by symmetry to facets other than test questions, the successive steps of a G-study, their results and their possible use to improve the reliability of measures, in line with the initial publications of Cardinet, Tourneur and Allal.

PALAVRAS-CHAVE: Generalizabilidade, ANOVA, modelo, medida, dispositivo, fidelidade, variância de erro, optimização

Este artigo apresenta a noção de generalizabilidade, a sua origem nos trabalhos psicométricos de Cronbach, a sua extensão por simetria a diversos domínios para além do teste de questões, os vários passos de um estudo-G, os resultados directos destes e a sua possível utilização para melhorar a fidelidade das medidas, na perspectiva das publicações iniciais de Cardinet, Tourneur e Allal.

La théorie de la généralisabilité est une théorie de la mesure mise au point par Cronbach et ses collaborateurs dans le but initial d'unifier les différentes définitions et formules de la fidélité élaborées en psychométrie classique. Elle se réfère à l'analyse de la variance pour estimer et améliorer la précision des plans de mesure.

Dans cet article, nous donnons un bref aperçu de cette théorie en nous appuyant, autant que possible, sur l'exemple fictif de la recherche ci-après : « Dans un mémoire de licence en sciences de l'éducation, un étudiant s'est proposé d'analyser la relation entre la langue maternelle et les performances en calcul mental des adultes analphabètes. Il a administré, pour cela, une épreuve de cinq questions à 20 adultes choisis aléatoirement dans deux groupes linguistiques, à savoir le groupe bantou et le groupe non bantou. »

Notion de généralisabilité

Pour effectuer une mesure, il faut inévitablement opérer un certain nombre de choix. Ainsi, mesurer les performances des analphabètes en calcul mental a nécessité, pour l'exemple présenté ci-avant, le recours à une épreuve composée de quelques questions (5) et la sélection de quelques sujets (10 par groupe linguistique).

Toutefois, le résultat obtenu dans ces conditions réduites de l'observation n'est intéressant que dans la mesure où il constitue une bonne estimation du score qu'on obtiendrait à l'ensemble illimité des questions que l'on peut poser en calcul mental et pour la population de tous les adultes analphabètes bantous et non bantous. Une mesure n'est donc valable que si elle est généralisable à l'univers des conditions d'observation.

Par généralisabilité, il faut entendre le degré auquel on peut généraliser d'une observation particulière à la valeur recherchée, cette dernière étant non observable. Cette qualité englobe essentiellement trois caractéristiques classiques de la mesure : la fidélité, la justesse et la sensibilité.

Genèse de la théorie de la généralisabilité

Le modèle de généralisabilité est l'extension de la théorie classique de la fidélité. Ses origines lointaines se situent par conséquent dans les travaux réalisés par Spearman au début du XX^{e} siècle. On peut, à cet effet, se rappeler qu'en se fondant sur le fait que le score observé (X) est obtenu en ajoutant au score vrai (V) l'erreur de mesure (E), Spearman définit la fidélité comme étant le rapport de la variance des scores vrais à la variance des scores observés des personnes examinées, soit $\sigma^2_{(V)} / \sigma^2_{(X)}$ ou encore $\sigma^2_{(V)} / (\sigma^2_{(V)} + \sigma^2_{(E)})$.

Le score vrai d'un sujet est défini comme la moyenne des scores observés qu'il obtiendrait sous toutes les conditions possibles d'observation. L'erreur de mesure devient alors la différence entre le score observé et le score vrai

pour une de ces conditions. L'erreur n'a pas de ce fait une signification précise, elle est simplement considérée comme une sorte de perturbation de la mesure.

Par ailleurs, dans le modèle de Spearman, le concept de fidélité semble être ambigu puisque « la théorie classique ne considère qu'une seule source d'erreur à la fois : soit celle relative à l'homogénéité des items (on parle de consistance interne) ou bien celle relative à la consistance des mesures d'une administration à l'autre de l'instrument (c'est la stabilité) ou encore celle relative à la consistance des mesures d'une forme de l'instrument à l'autre (il s'agit de l'équivalence) ». (Bertrand, s.d., pp. 2-3.) La notion de fidélité prend ainsi des significations différentes et diverses procédures ont été établies pour sa mesure[1]. À ce propos, Cronbach et ses collaborateurs (1972, p. 5) notent que l'on peut obtenir des résultats contradictoires si on apprécie la fidélité des mesures selon ces différentes procédures.

Conscients de l'ambiguïté du concept de fidélité et des contradictions éventuelles auxquelles son appréciation par diverses méthodes peut donner lieu, Cronbach et ses collaborateurs se sont fixé l'objectif principal d'unifier le champ de la psychométrie. Pour cela, tout en admettant, à l'instar de Spearman, que le score observé est la somme du score vrai et de l'erreur de mesure, ils ont non seulement élaboré un modèle qui prend en compte plusieurs sources d'erreur à la fois, mais ils ont aussi donné une signification précise au concept d'erreur de mesure : *« fluctuation d'échantillonnage des conditions d'observation »*. Ainsi, par exemple, le score obtenu par un élève à une épreuve varie selon les questions posées, les correcteurs choisis, les moments d'évaluation, etc. Autrement dit, ce score (observé) s'écarte de la moyenne que l'élève obtiendrait (score vrai) à l'univers des questions, des correcteurs, des moments, etc., à cause du choix des conditions d'observation.

Cependant, bien qu'il soit un progrès important pour l'étude et l'amélioration de la fidélité par rapport à la psychométrie classique, le modèle de Cronbach a été exploité, autant que la théorie de Spearman, uniquement pour la différenciation des personnes. Il a fallu attendre l'exploitation du principe de symétrie[2] par Cardinet, Tourneur et Allal (1976, 1981) pour que l'application de la théorie de la généralisabilité soit étendue à d'autres objets d'étude tels que le classement des taux de réussite des objectifs d'apprentissage, la différenciation des cotations des correcteurs, la discrimination des phases d'apprentissage, etc. Ainsi établi, le modèle de généralisabilité peut être exploité comme une véritable théorie de la mesure en éducation.

Schéma d'une étude de généralisabilité

Le but d'une étude de généralisabilité est d'estimer l'importance des composantes de variance qui influent sur les observations afin de suggérer une procédure optimale permettant la réduction de l'erreur et une meilleure généralisation des résultats. Son déroulement s'opère en deux grandes phases : l'analyse de la variance et l'étude de généralisabilité proprement dite.

Analyse de la variance

Cette première phase vise à découvrir et à quantifier, par l'analyse de la variance, les sources de variation qui affectent la mesure. Elle comprend le plan d'observation et le plan d'estimation.

Plan d'observation

C'est la structuration ou la planification des données en fonction des conditions sous lesquelles elles sont mesurées. Plus concrètement, la définition du plan d'observation exige la recherche des différentes variables ou facettes qui sont à la base de la variation observée des données, la détermination du nombre de niveaux de chaque facteur et l'explicitation des relations entre les facettes.

À propos de l'exemple mentionné plus haut, le plan d'observation comprend trois facettes : les *Sujets (S),* les *Groupes linguistiques (G)* et les *Questions (Q)*. Le nombre de niveaux de ces facettes est respectivement de dix pour les sujets, deux pour les groupes linguistiques et cinq pour les questions. Le nombre total d'observations pour le plan complet, donné par le produit des nombres de tous ces niveaux, est de 100 (soit 10 x 2 x 5). Alors qu'ils sont nichés dans les groupes linguistiques, les sujets sont croisés avec les questions. C'est également le croisement qui relie les groupes linguistiques aux questions.

Deux facettes sont dites croisées lorsque, pour chaque niveau d'une facette, il y a des résultats pour chaque niveau de l'autre facette. Noté par le signe de multiplication (x) placé entre les facettes (exemple S x Q), le croisement est illustré par des diagrammes qui s'entrecoupent (*cf*. figure 1).

Le nichage est une relation d'emboîtement ou d'inclusion. Symbolisé par un double point (:) placé entre la facette nichée (à gauche) et la facette nichante (à droite), le nichage est graphiquement traduit par l'inclusion du diagramme de la facette nichée dans celui de la facette nichante (*cf*. figure 1).

Compte tenu des relations entre les facettes S, G et Q, le plan d'observation de la recherche mentionnée au seuil de ce texte peut être graphiquement illustré par le diagramme de la figure 1.

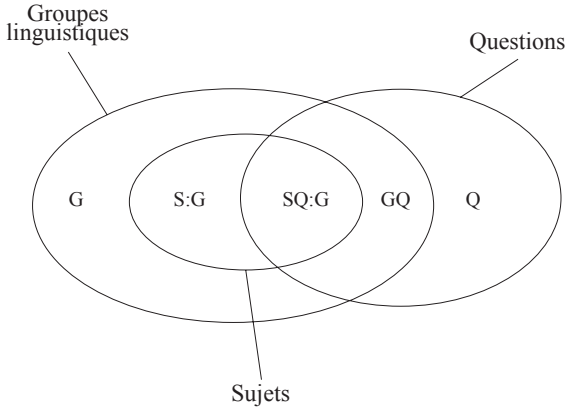

Figure 1. ***Plan d'observation (S_{10} : G_2) x Q_5***

Ce plan d'observation comprend cinq sources de variation : trois effets principaux (G, S:G et Q) et deux effets d'interaction (GQ et SQ :G).

Plan d'estimation

Comme déjà indiqué, les résultats réalisés par 20 sujets analphabètes aux cinq questions de l'épreuve ne sont pas intéressants en eux-mêmes, mais seulement dans la mesure où ils permettent d'estimer ce qu'on aurait obtenu avec l'ensemble de la population des analphabètes bantous et non bantous confrontés à l'infinité des questions de calcul mental. Il faut donc aller au-delà du plan d'observation, qui se limite uniquement aux données échantillonnées, pour calculer les carrés moyens, et estimer la variance qu'on obtiendrait pour chaque effet dans le cadre de la population ou de l'univers de référence[3]. Il s'agit de décomposer la variance totale en autant de composantes qu'il y a de sources de variation dans le plan d'analyse.

Cette décomposition exige que l'on précise au préalable la façon dont les niveaux de chaque facette ont été sélectionnés. Ces niveaux ont-ils été pris dans leur totalité ou ont-ils été choisis aléatoirement ? Par exemple, pour l'exemple analysé dans ce texte, alors que les groupes linguistiques sont une facette fixée, les sujets et les questions sont échantillonnés, ou considérés comme échantillonnés, à partir d'univers infinis.

Ces renseignements permettent d'estimer les composantes de variance à partir des carrés moyens. Pour le faire, les algorithmes de calcul sont plus simples si l'on procède en deux étapes. On considère d'abord que les niveaux de chaque facette du plan d'observation ont été extraits aléatoirement à partir d'un univers infini. Les composantes de variance ainsi calculées, appelées

composantes de variance aléatoires, sont ensuite transformées en composantes de variance mixtes. Cette conversion tient compte de la manière dont les niveaux de chaque facette retenue dans le plan d'observation ont effectivement été sélectionnés.

Les composantes de variance mixtes ne sont cependant pas additives. En d'autres termes, « si on les additionne, on n'obtient pas la meilleure estimation de la variance totale des résultats observés. Pour les rendre additives, une correction est nécessaire (dite correction de Whimbey[4]). » (Bain & Pini, 1996, p. 21.) Les composantes de variance mixtes corrigées sont appelées espérances de variance mixtes.

En appliquant aux données reprises en annexe le logiciel des études de généralisabilité (GT) mis au point par Ysewijn (1996), nous obtenons les résultats ci-dessous de la première phase de l'étude de généralisabilité, c'est-à-dire l'analyse de la variance.

Tableau 1
Analyse de la variance pour le plan (S : G) x Q

Plan d'observation				*Plan d'estimation*				
Sources de variation	*Sommes/ carrés*	*Deg./ liberté*	*Carrés moyens*	*Composantes aléatoires*	*Composantes mixtes*	*Espérances variance mixtes*	*Erreurs types*	*% de variance*
G	2,56	1	2,560	0,04022	0,04022	0,02011	0.01196	8
S:G	9,68	18	0,538	0,08778	0,08778	0,08778	0.03417	33
Q	4,84	4	1,210	0,05500	0,05556	0,05556	0.03508	21
GQ	0,44	4	0,110	0,00111	0,00111	0,00056	0.00656	0
SQ:G	7,12	72	0,099	0,09889	0,09889	0,09889	0.01626	38
Totaux	24,64	99		0,28300	0,28356	0,26289		100

L'observation de la dernière colonne de ce tableau indique que c'est l'interaction *Sujets* x *Questions* (SQ : G) qui contribue le plus à la variation totale. Elle représente 38 % de cette dernière. Cette source de variation est suivie des effets principaux *Sujets* (S:G) et *Questions* (Q) dont les contributions à la variance totale sont respectivement de 33 % et 21 %. En revanche, l'effet principal *Groupes* (G) ne contribue que faiblement à la variance totale des scores observés (8 %) tandis que l'interaction entre les groupes linguistiques et les questions (GQ) ne concourt pas du tout à cette variation.

Par ailleurs, les erreurs types (avant-dernière colonne) permettent d'apprécier l'effet des fluctuations d'échantillonnage sur le calcul des composantes de variance aléatoires. Pour être significative, chaque composante aléatoire doit être plus grande que deux erreurs-types. Au regard de ce critère, seulement deux composantes de variance aléatoires sont significatives (soit $\sigma^2_{S:G}$ et $\sigma^2_{SQ:G}$). On peut déjà anticiper que le dispositif ne permet pas de différencier les niveaux des autres sources de variation dont les composantes aléatoires ne sont pas significativement différentes de 0. Mais pour plus de précision, il faut procéder à l'étude de généralisabilité proprement dite.

Étude de généralisabilité proprement dite

À cette seconde phase de l'étude de généralisabilité, on spécifie d'abord le rôle de chaque facette dans la mesure, ce qui permet d'apprécier la précision du plan de mesure, et on suggère ensuite des aménagements susceptibles de réduire l'erreur de la mesure. Ces opérations sont effectuées à travers le plan de mesure et le plan d'optimisation.

Plan de mesure

Au cours de deux étapes antérieures, les facettes sont, en vertu du principe de symétrie qui régit l'analyse de la variance, traitées de la même façon. Par contre, le plan de mesure introduit un *distinguo* entre les facettes en définissant, d'une part, les objets d'étude (c'est-à-dire les facettes qui sont la cible d'évaluation) et, d'autre part, les instruments (c'est-à-dire les facettes qui sont utilisées comme moyens) de cette évaluation. Les objets d'étude constituent ce qu'on appelle *face de différenciation* tandis que les conditions d'observation forment la *face d'instrumentation*. Aussi, dans la symbolisation du plan de mesure, les facettes de différenciation sont placées à gauche de la barre de fraction tandis que les facettes d'instrumentation sont classées à droite, soit : Face de différenciation/Face d'instrumentation[5].

Le placement des facettes sur les faces du plan de mesure permet d'estimer les paramètres de généralisabilité (la variance de différenciation, la variance d'erreur relative et la variance d'erreur absolue) ainsi que les coefficients de généralisabilité ou coefficients d'assurance.

Appelée également variance univers, la variance de différenciation permet d'apprécier la différenciation des objets d'étude. Elle est obtenue par la somme des composantes de variance associées aux facettes de différenciation et à leurs interactions. La variance d'erreur provient quant à elle des composantes de variance associées aux facettes d'instrumentation aléatoires

(ou facettes de généralisation). Les facettes d'instrumentation fixées (ou facettes de contrôle) ne sont pas prises en compte pour la détermination de la variance d'erreur, puisqu'elles ne créent pas de fluctuations d'échantillonnage.

Suivant le type de mesure envisagé (mesure absolue ou relative)[6], on calcule soit la variance d'erreur absolue, soit la variance d'erreur relative. «La variance d'erreur absolue s'obtient en additionnant toutes les composantes de variance du plan qui n'interviennent pas dans la variance de différenciation, chaque composante étant divisée par le nombre de niveaux des facettes de généralisation qui apparaissent dans son indice total[7].» (Duquesne, 1987, p.52.) La variance d'erreur relative s'obtient en additionnant, avec leurs coefficients, toutes les composantes de la variance absolue qui comprennent au moins une facette de différenciation dans leur indice total.

À partir des paramètres de généralisabilité, il est aisé de calculer le coefficient de généralisabilité (une mesure de la fidélité), c'est-à-dire le rapport entre la variance de différenciation et la somme de la variance de différenciation et la variance d'erreur. Sa formule est la suivante :

$$E\rho^2 = \sigma^2_{(\tau)} / \sigma^2_{(\tau)} + \sigma^2_{(e)}$$

dans laquelle :

$E\rho^2$ = coefficient de généralisabilité

$\sigma^2_{(\tau)}$ = variance univers (de différenciation)

$\sigma^2_{(e)}$ = variance d'erreur

Pour chaque plan de mesure, deux coefficients de généralisabilité peuvent être calculés, selon qu'on s'intéresse à la mesure absolue ou à la mesure relative. Le coefficient de généralisabilité absolu ($E\rho^2_{\Delta}$) est calculé avec la variance d'erreur absolue ($\sigma^2_{(\Delta)}$) tandis que le coefficient de généralisabilité relatif ($E\rho^2_{\delta}$) l'est avec la variance d'erreur relative ($\sigma^2_{(\delta)}$).

Variant de 0 à 1, ce coefficient indique une fidélité insuffisante lorsqu'il tend vers 0 et une fidélité suffisante lorsqu'il se rapproche de 1. Habituellement, on admet que la fidélité est suffisante lorsque la valeur de $E\rho^2$ est supérieure ou égale à 0,80, c'est-à-dire lorsque la variance de différenciation est quatre fois plus importante que la variance d'erreur.

Pour le mémoire de licence relevé plus haut, l'objectif poursuivi étant la comparaison des performances de deux groupes d'analphabètes en calcul mental, la facette *Groupes linguistiques* (G) constitue par conséquent l'objet d'étude et appartient à la face de différenciation.

La comparaison des groupes linguistiques est assurée au moyen des résultats obtenus par les analphabètes aux cinq questions de l'épreuve de calcul mental. Dès lors, les facettes *Sujets* (S) et *Questions* (Q) sont des conditions d'observation ; autrement dit, elles appartiennent à la face d'instrumentation. S et Q sont des facettes d'instrumentation aléatoires ou facettes de généralisation, étant donné que les 20 analphabètes retenus dans l'enquête et les cinq questions couvertes par l'épreuve ont été choisis (ou considérés comme choisis) aléatoirement dans des univers infinis.

Ce plan de mesure, dont la représentation graphique apparaît à la figure 2, peut être symbolisé par G/SQ. Les hachures de la face de différenciation sont horizontales ; celles de la face d'instrumentation, verticales.

La variance de différenciation (variance vraie ou variance univers) pour le plan de mesure G/SQ est estimée par l'espérance de variance *Groupes* ($\sigma^2_{(G)}$), soit 0,02011 (*cf.* tableau 2).

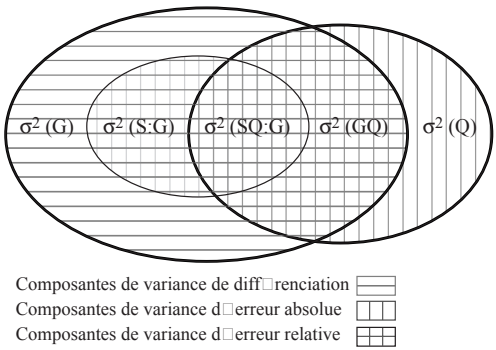

Figure 2. ***Plan de mesure G/SQ***

La variance d'erreur absolue est la somme des quatre autres composantes de variances ($\sigma^2_{(S:G)}$, $\sigma^2_{(SQ:G)}$, $\sigma^2_{(GQ)}$, $\sigma^2_{(Q)}$), chaque composante étant divisée par le nombre de niveaux des facettes d'instrumentation aléatoires (S, Q) apparaissant en indice primaire. Pour obtenir la variance d'erreur relative, il suffit d'additionner à partir de cette variance d'erreur absolue uniquement les composantes comprenant au moins la facette de différenciation (G) dans leur indice total, soit, $\sigma^2_{(S:G)}$, $\sigma^2_{(SQ:G)}$, et $\sigma^2_{(GQ)}$.

Pour le plan de mesure G/SQ, les résultats de l'analyse des paramètres de généralisabilité sont synthétisés dans le tableau 2.

Tableau 2
Analyse de généralisabilité du plan de mesure G/SQ

Sources de variation	*Variance de différenciation*	*Sources de variation*	*Variance d'erreur relative*	*Variance d'erreur absolue*
G	0,02011			
		S:G	0,00878	0,00878
		Q		0,01111
		GQ	0,00011	0,00011
		SQ:G	0,00198	0,00198
Totaux	0,02011		0,01087	0,02198

Coefficients de généralisabilité relatif 0,649, absolu 0,478.

Les coefficients de généralisabilité relatif (0,649) et absolu (0,478) étant de loin inférieurs à la limite généralement considérée comme satisfaisante (0,80), la précision de la mesure fournie par le plan d'observation (S:G) x Q est faible.

Dans les deux cas de la mesure (relative et absolue), la variance d'erreur est très importante par rapport à la variance de différenciation. Le dispositif de mesure n'est donc pas assez puissant pour assurer la comparaison des deux groupes linguistiques selon leurs performances en calcul mental (mesure relative), ni la comparaison des performances de chaque groupe linguistique par rapport à une norme donnée (mesure absolue).

Plan d'optimisation

L'amélioration de la fidélité s'opère à la quatrième étape d'une étude de généralisabilité, c'est-à-dire le plan d'optimisation (appelé également étude D ou étude de décision). Cette optimisation résulte des interventions que l'on effectue sur les plans d'observation, d'estimation et de mesure.

Bien que les pistes d'optimisation préconisées dans la théorie de la généralisabilité soient nombreuses, on peut en distinguer quatre principales : la modification du nombre de niveaux échantillonnés, l'élimination des niveaux atypiques par l'analyse de facettes[8], la modification du regroupement et du mode d'échantillonnage des facettes et l'amélioration de la validité du dispositif par l'élimination des biais de mesure[9].

Pour assurer la comparaison des groupes linguistiques, nous illustrons ci-après deux de ces directions, à savoir la modification du nombre de niveaux échantillonnés et la modification du mode d'échantillonnage des facettes.

Mais sur quelles conditions d'observation (sur quelles facettes) faut-il agir pour améliorer la généralisabilité des mesures ?

Pour répondre à cette question, il convient d'établir, à partir des paramètres de généralisabilité (*cf.* tableau 2), les sources de variation qui contribuent le plus à la variance d'erreur.

Cet examen indique que c'est S:G qui concourt le plus à la variance d'erreur relative. Sa part (0,00878) représente 81 % de la variance d'erreur relative (0,01087). Pour la maximisation de la mesure absolue, il faut agir simultanément sur les facettes *Questions* et *Sujets* dont les contributions (0,01111 et 0,00878) représentent respectivement 50 % et 40 % de la variance d'erreur absolue.

Modification du nombre de niveaux échantillonnés. Quelques propositions d'augmentation du nombre de niveaux des facettes *Sujets* et *Questions* sont présentées dans le tableau 3.

Tableau 3
Optimisation du plan de mesure G/SQ par l'augmentation des nombres de niveaux des facettes
Sujets *et* Questions

Facette	*G*	*S:G*	*Q*	*Nombre d'observations*	*Coeff. de généralis. relatif*	*Coeff. de généralis. absolu*
Niveaux pour le plan de base	2	10	5	100	0,649	0,478
Optimisation 1	2	10	10	200	0,672	0,567
Optimisation 2	2	15	5	150	0,734	0,522
Optimisation 3	2	20	10	400	0,803	0,657

Doubler le nombre de questions sans modifier le nombre de sujets (optimisation 1) n'améliore sensiblement ni la mesure relative, ni la mesure absolue. Mais, si l'on choisit 15 sujets par groupe linguistique sans modifier la longueur de l'épreuve de calcul mental (optimisation 2), la précision de la mesure s'améliore un peu plus notablement. Cette amélioration est plus importante encore si l'on double à la fois la taille de l'échantillon des sujets et la longueur de l'épreuve (optimisation 3). Cette dernière solution permet au coefficient de généralisabilité relatif d'atteindre la limite de 0,80.

Modification du mode d'échantillonnage des facettes. En considérant par exemple la facette Sujets comme une facette d'instrumentation fixée, en renonçant par conséquent à étendre la différenciation des groupes linguistiques à tous les adultes analphabètes bantous et non bantous, on accroît considérablement la précision de la mesure relative. Cette solution fait passer le coefficient de généralisabilité relatif de 0,649 à 0,957. La même modification du mode d'échantillonnage de la facette Sujets augmente notoirement le coefficient de généralisabilité absolu (de 0,478 à 0,650).

Discussion de ces modifications. Réduire la part d'erreur d'échantillonnage dans la mesure est donc possible, mais au prix d'autres inconvénients, parfois rédhibitoires.

Augmenter le nombre de niveaux d'une facette de généralisation oblige d'abord à reprendre d'autres observations, ce qui peut être onéreux en temps et en moyens matériels. Il faut veiller aussi à ce que le nouvel échantillon reste représentatif de la même population.

Fixer une facette d'instrumentation réduit considérablement l'intérêt de la mesure. Dans notre exemple, fixer la facette *Questions* signifierait qu'on ne pourrait plus parler de la capacité de « calculer mentalement », mais seulement du savoir effectuer ces cinq opérations particulières. Fixer la facette *Sujets* ôterait toute portée aux résultats.

Quelques exigences d'une étude de généralisabilité

Deux grandes exigences, dues au modèle d'ANOVA exploité par la théorie de la généralisabilité, sont à respecter lorsqu'on effectue une étude de généralisabilité : l'une a trait à la nature des plans d'observation et l'autre se rapporte à l'échelle d'évaluation à laquelle on se réfère. En effet, les logiciels des études de généralisabilité actuellement disponibles (notamment Etudgen, GT, Win-GT, EduG) ne sont applicables qu'à des plans d'observation équilibrés, c'est-à-dire des plans qui comportent un même nombre d'éléments de la facette nichée à chaque niveau de la facette nichante. Pour l'exemple exploité dans ce texte, puisque la facette *Sujets* est nichée dans la facette *Groupes linguistiques*, nous avons été contraint de travailler avec autant de sujets bantous que de sujets non bantous (soit dix sujets par groupe linguistique).

Les logiciels des études de généralisabilité exigent en outre que « les valeurs sur lesquelles on travaille se réfèrent à une même échelle : même maximum ou minimum. Ainsi, tous les items d'une épreuve doivent être cotés

sur le même nombre de points. Si ce n'est pas le cas, on recourra à des transformations pour ramener les items à une même échelle» (Bain & Pini, 1996, p. 8).

Annexe

Résultats (fictifs) d'analphabètes bantous et non bantous à l'épreuve de calcul mental

Groupes linguistiques	*Sujets*	Q_1	Q_2	Q_3	Q_4	Q_5
Bantou	1	0	0	0	0	0
	2	0	0	0	0	0
	3	0	0	0	0	0
	4	0	0	0	0	1
	5	0	0	0	0	1
	6	0	0	1	1	1
	7	0	0	1	1	1
	8	0	0	1	1	1
	9	0	1	1	1	1
	10	1	1	1	1	1
Non bantou	11	1	1	1	1	1
	12	0	0	0	1	1
	13	0	0	0	1	1
	14	0	0	0	1	1
	15	0	0	0	1	1
	16	0	1	1	1	1
	17	0	1	1	1	1
	18	1	1	1	1	1
	19	1	1	1	1	1
	20	1	1	1	1	1

NOTES

1. Les principales procédures de l'étude de la fidélité élaborées en psychométrie classique sont notamment le test-retest, les formes parallèles, la méthode de la bipartition et la consistance interitem.
2. Le principe de symétrie est «la propriété du modèle d'analyse de la variance qui consiste à ne pas privilégier un facteur par rapport à d'autres, toutes les sources de variation contribuant parallèlement à la variance totale» (Cardinet & Tourneur, 1985, p. 328).
3. Cette variance est appelée composante de variance.

4. Cette correction consiste à multiplier chaque composante mixte par (N-1)/N, N étant le nombre de niveaux de l'univers de chaque facette apparaissant en indice primaire de l'effet, c'est-à-dire à gauche du signe de nichage (:).
5. Si nous notons G/SQ, pour l'exemple étudié dans ce texte, cela revient à dire que nous visons la différenciation des groupes linguistiques (G) au moyen des sujets (S) et des questions (Q).
6. « On parle de mesure absolue lorsqu'on cherche à situer une grandeur par rapport à une échelle dont les échelons sont définis *a priori*. C'est par exemple le cas lorsqu'on s'intéresse aux scores en termes de nombres de points obtenus par les personnes aux différents items. Par contre, on parle de la mesure relative lorsque ce sont les positions relatives des résultats qui constituent l'information essentielle, comme par exemple lorsqu'on s'intéresse aux scores, non pour leurs quantités respectives, mais pour l'écart qui existe entre eux. » (Duquesne, 1988-1989, p. 187.)
7. L'indice total comprend toutes les facettes apparaissant aussi bien avant qu'après le signe du nichage. Par exemple pour la composante de variance $\sigma^2_{SQ:G}$ l'indice primaire est composé de S et Q tandis que l'indice total comprend S, Q et G.
8. L'analyse de facettes est l'équivalent de l'analyse classique des items (lire à ce sujet Mokonzi, *Méthodes d'analyse d'items et optimisation de la fiabilité des mesures en éducation*, texte publié dans ce même numéro de la revue).
9. Pour plus de renseignements sur ces directions d'optimisation de la mesure, lire Bain et Pini (1996).

RÉFÉRENCES

Bain, D. & Pini, G. (1996). *Pour évaluer vos évaluations. La généralisabilité : mode d'emploi*. Genève : Centre de recherches psychopédagogiques. Direction générale du cycle d'orientation.

Bertrand, R. (s. d.). *Pourquoi de nouvelles théories de la mesure ?* Sainte-Foy : Université Laval.

Cardinet, J., Tourneur, Y. & Allal, L. (1976). The symmetry of generalizability theory : applications to educational measurement. *Journal of Educational Measurement, 13*, 119-135.

Cardinet, J., Tourneur, Y. & Allal, L. (1981). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement*, 18, 183-204.

Cardinet, J. & Tourneur, Y. (1985). *Assurer la mesure*. Berne : Peter Lang.

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements : theory of generalisability for scores and profiles*. New York : J. Wiley.

Duquesne, F. (1987). Le calcul des paramètres de généralisabilité. *Informatique et sciences humaines, 72-73*, 41-78.

Duquesne, F. (1988-1989). La théorie de la généralisabilité : un modèle pour assurer la différenciation des objets d'étude. *Bulletin de psychologie, XLII 388*, 185-189.

Ysewijn, P. (1996). *G.T. : Logiciel pour les études de généralisabilité, version 1.61 F*. Document inédit : Bercher.