

La fidélité des scores totaux et la fidélité des scores logits : le cas du modèle de Rasch

Reliability of Total Scores and Logit Scores: the Case of the Rasch Model

Sébastien Béland – sebastien.beland@umontreal.ca – <https://orcid.org/0000-0002-0324-5076>

Université de Montréal - Canada

Christian Bourassa – christian.bourassa.1@umontreal.ca – <https://orcid.org/0000-0002-6648-498X>

Université de Montréal - Canada

Eric Dionne - Eric.Dionne@uOttawa.ca – <https://orcid.org/0000-0003-3046-4386>

Université d'Ottawa - Canada

Pour citer cet article : Béland, S., Bourassa, C. et Dionne, E. (2024). La fidélité des scores totaux et la fidélité des scores logits : le cas du modèle de Rasch. *Évaluer. Journal international de recherche en éducation et formation*, 10(2), 71-84. <https://doi.org/10.48782/e-jiref-10-2-71>

Résumé

Nous sommes nombreux à être confondus avec les perspectives de la fidélité lors d'une analyse à l'aide du modèle de Rasch pour réponses dichotomiques. Ce court article vise à illustrer la différence existante entre la fidélité des scores totaux et la fidélité des scores exprimés sous la forme d'un logit. Pour ce faire, nous baserons notre argumentation sur deux stratégies qui quantifient la fidélité d'un ensemble de données dichotomiques unidimensionnelles dans le contexte du modèle de Rasch : la stratégie développée par Dimitrov et l'indice de séparation des personnes. Contrairement à ce que certains prétendent, les deux approches ne sont pas interchangeables. Nous allons aussi faire quelques constats pour la recherche.

Mots-clés

Fidélité des scores totaux, fidélité des scores logits, théorie classique des tests, modèle de Rasch

Abstract

Many of us find the concept of reliability confusing when applying the Rasch model to dichotomous responses. This article aims to clarify the distinction between the reliability of total scores and the reliability of logit scores. To do so, we will focus on two approaches that quantify the reliability of one-dimensional dichotomous data within the Rasch model: Dimitrov's strategy and the Person Separation Index. Contrary to what some colleagues suggest, these two methods are not interchangeable. We will also offer some observations relevant to future research.

Keywords

Reliability of total scores, reliability of logit scores, classical test theory, Rasch model

1. Introduction

Le contenu de cet article n'est pas nouveau, mais il a l'avantage de clarifier une différence importante entre la théorie classique des tests (TCT) et le modèle de Rasch/la théorie de la réponse aux items (pour une discussion plus étendue voir Dimitrov, 2023). Celui-ci s'insère aussi dans le mouvement plus large faisant la promotion des bonnes pratiques d'analyses quantitatives en sciences sociales (Finkel *et al.*, 2015; Sijtsma, 2016).

Plusieurs chercheurs sous-estiment le défi de bien comprendre la différence entre le concept de fidélité des scores totaux, basée sur la TCT, et la fidélité des scores logits, basée sur la modélisation de Rasch pour réponses dichotomiques. En effet, Wright (1996, p. 472) a établi, il y a longtemps, la nuance entre la variance « vraie » qui est associée à l'échantillon des sujets qui complètent le test et la variance de l'erreur qui, elle, est une caractéristique de l'instrument de mesure. Cette simple nuance est souvent ignorée dans les écrits scientifiques alors que les auteurs se retranchent dans une conception de la fidélité purement arithmétique. Cela n'est pas surprenant puisque plusieurs ouvrages tablent essentiellement sur la ressemblance entre ces deux conceptions de la fidélité. Un exemple éloquent est l'ouvrage classique de Wright et Stone (1999, p. 153) qui présente seulement leurs similitudes. De leur côté, Andrich et Marais (2019, p. 129) déclarent qu'il est possible de calculer la fidélité à partir du modèle de Rasch à l'aide du coefficient de séparation de personne (ISP), mais la distinction reste essentiellement au niveau des équations mathématiques. Or, des auteurs tels que Bartholomew *et al.* (1993) ont parlé de « fidélité du trait latent »¹, qui est basée sur les scores logits. Les écrits ne sont toutefois pas toujours clairs pour distinguer ces deux types de fidélité.

Au moment d'écrire ces lignes, la revue qui s'intéresse spécifiquement au développement du modèle de Rasch, le *Journal of Applied Measurement*, proposait de ne pas rapporter le coefficient alpha (α) de Cronbach, lié à la première approche, pour plutôt rapporter l'ISP, lié à la deuxième approche. Lors de nos nombreuses interventions dans différents milieux de recherche, nous avons souvent entendu que la fidélité des scores totaux était incompatible avec une analyse à l'aide du modèle de Rasch. Or, nous croyons que cela est inexact.

Nous nous positionnons, ici, en expliquant qu'il est possible de mobiliser une approche de la fidélité basée sur les scores totaux en utilisant le coefficient proposé par Dimitrov (2003ab), et celle-ci n'est pas la même chose que l'ISP, qui s'insère dans la logique de la fidélité des scores logits.

Cet article poursuit deux objectifs. Premièrement, nous souhaitons illustrer la différence entre la fidélité des scores totaux et la fidélité des scores logits à partir du modèle de Rasch et, ce, à l'aide de données simulées. À notre connaissance, cet exercice n'a pas été encore fait. Ensuite, nous voulons formuler des recommandations pour mieux rapporter la fidélité dans le cadre d'une analyse avec le modèle de Rasch. Nous croyons que les chercheurs et les praticiens qui utilisent cette modélisation doivent être vigilants au moment de choisir une stratégie ou l'autre de la fidélité, car elles ne sont pas substituables et peuvent entraîner des décisions différentes en ce qui concerne, par exemple, le retrait de certains items. Ajoutons également que la valeur du coefficient α de Cronbach souffre de nombreux problèmes, dont

¹ Dans le cadre de cet article, nous emploierons l'expression « fidélité des scores logits », qui nous semble plus appropriée.

celui d'un effet plancher en raison de la non-linéarité des scores totaux (Mallinson *et al.*, 2004) et de l'identification de la borne inférieure de la fidélité (Waller et Revelle, 2023).

2. La fidélité dans le cadre de la TCT

Nous commencerons par examiner la conception qui est la plus connue, soit celle dans le cadre de la TCT (Lord et Novick, 1968) : la fidélité des scores totaux. Bien que plusieurs stratégies existent (par exemple, le test-retest et les versions parallèles), nous allons nous focaliser sur la situation où les réponses proviennent d'une seule passation.

Selon cette théorie, le score observé Y d'un répondant à un test ou à une échelle de mesure correspond à la somme de deux composantes : le score vrai V et l'erreur de mesure aléatoire E , donc $Y=V+E$. Quand les observations d'un groupe de répondants sont disponibles, alors la variance du score observé Y peut être décomposée de façon similaire, soit $\sigma_Y^2 = \sigma_V^2 + \sigma_E^2$.

La corrélation (notée ρ , par convention) entre les scores observés Y d'un groupe de répondants et les scores vrais V de ces mêmes individus prend la forme suivante :

$$\rho(Y, V) = \frac{\sigma_{Y,V}}{\sigma_Y \sigma_V}$$

où $\sigma_{Y,V}$ est la covariance entre Y et V , alors que $\sigma_Y \sigma_V$ est la multiplication de l'écart type de Y avec celui de V . Ce coefficient devrait idéalement être élevé, soit près de l'unité (en valeur absolue), pour un test présentant une bonne fidélité. Traditionnellement, le carré de ρ est assimilable à un coefficient de détermination et il est utilisé pour quantifier la fidélité des scores totaux, ainsi :

$$\rho(Y, V)^2 = \rho_{TCT}^2 = \left(\frac{\sigma_{Y,V}}{\sigma_Y \sigma_V} \right)^2 = \left(\frac{\sigma_V^2}{\sigma_Y \sigma_V} \right)^2 = \left(\frac{\sigma_V}{\sigma_Y} \right)^2 = \frac{\sigma_V^2 - \sigma_E^2}{\sigma_V^2 + \sigma_E^2} = 1 - \frac{\sigma_E^2}{\sigma_V^2} \quad (2)$$

et les deux derniers termes de cette équation correspondent à la définition de la fidélité telle qu'elle figure dans plusieurs manuels d'introduction à la mesure en éducation. Les valeurs possibles de ρ_{TCT}^2 sont donc nécessairement bornées entre zéro et un². Notons qu'il existe une autre définition basée sur des tests parallèles, mais nous n'en parlerons pas pour simplifier notre argumentation (voir Sijtsma et Pfadt (2021) pour un exemple récent).

² Il est intéressant de savoir que le coefficient α , qui estime ρ_{TCT}^2 , peut prendre des valeurs négatives (Martineau, 1982).

3. La fidélité dans le cadre du modèle de Rasch

Supposons que nous avons obtenu les réponses dichotomiques x (0 ou 1) de J personnes à I items qui forment la matrice suivante :

Tableau 1 : Matrice de réponses

		Réponses aux items			
		$i=1$	$i=2$...	I
Personnes	$j=1$	x_{11}	x_{12}	...	x_{1I}
	$j=2$	x_{21}	x_{22}	...	x_{2I}

	J	x_{J1}	x_{J2}	...	x_{JI}

Le modèle de Rasch pour réponse dichotomique (Rasch, 1960; Dionne et Béland, 2023) peut être utilisé afin de calculer la probabilité qu’une personne obtienne une bonne réponse à un item. Mathématiquement, celui-ci prend la forme suivante :

$$P(x_{ji} = 1 | \theta_j, b_i) = P(\theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (3)$$

où x_{ij} correspond à la réponse dichotomique de la personne j à l’item i , b_i est la difficulté de l’item, et θ_j correspond au niveau d’habileté de la personne j . Ici, $P(\theta_j)$ est la probabilité d’obtenir une bonne réponse à un item. Ce modèle postule, entre autres, de l’unidimensionnalité du niveau d’habileté et de l’indépendance locale entre les items. Ces paramètres peuvent être estimés et devraient être documentés à l’aide de différentes méthodes comme l’estimation par maximum de vraisemblance conjointe ou par maximum de vraisemblance conditionnelle (voir Baker et Kim (2004) pour de plus amples détails).

L’idée derrière l’équation 2 peut être appliquée dans le contexte du modèle de Rasch, mais au lieu d’utiliser la décomposition $Y = V + E$, c’est plutôt la décomposition du niveau d’habileté estimé qui est mobilisée, soit $\hat{\theta}_j = \theta_j + E_j$ et qui nous permet de comprendre que $\theta_j = \hat{\theta}_j - E_j$ ³.

Pour Andrich et Marais (2019, p.130), la conception théorique de la fidélité est définie ainsi dans le cadre du modèle de Rasch :

$$\rho_{Rasch}^2 = \frac{\sigma_{\theta}^2 - \sigma_E^2}{\sigma_{\theta}^2} = 1 - \frac{\sigma_E^2}{\sigma_{\theta}^2} \quad (4)$$

qui peut se comprendre comme la part de la variance totale des scores logits σ_{θ}^2 après avoir enlevé l’erreur de mesure σ_E^2 . Le lecteur curieux pourra constater la similitude entre les équations 2 et 4, mais il doit garder en tête qu’elles ne correspondent pas exactement au même construit. Nous allons maintenant aborder deux stratégies pour estimer la fidélité dans le cadre du modèle de Rasch.

³ L’accent « ^ » est habituellement utilisé pour identifier une valeur estimée; à partir d’ici, il sera écarté pour simplifier la lecture de l’article.

4. Stratégies principales pour estimer la fidélité dans le cadre du modèle de Rasch

Examinons d'abord la procédure de Dimitrov et, ensuite, les aspects théoriques associés à l'ISP.

4.1. La procédure de Dimitrov pour estimer $\rho_{TCI}^2: D_{Rasch}$

Dimitrov (2003b) a publié une procédure analytique basée sur les scores totaux afin d'estimer la fidélité en s'appuyant sur les paramètres estimés à partir du modèle de Rasch pour réponse dichotomique. Pour cette raison, elle est en phase avec la définition de la fidélité présentée à l'équation 2. Ici, la variance d'erreur pour l'item i s'écrit :

$$\sigma_{E_i}^2 = \int_{-\infty}^{\infty} P(\theta_j)[1 - P(\theta_j)]\phi(\theta_j)d\theta_j \quad (5)$$

où la fonction de densité de l'intégrale des valeurs de la distribution latente est $\phi(\theta_j)$. L'erreur pour l'ensemble du test est donc la somme de l'erreur à chacun des I items, soit :

$$\sigma_E^2 = \sum_{i=1}^I \sigma_{E_i}^2. \quad (6)$$

La variance du score vrai V , pour un item, est mathématiquement égale à :

$$\sigma_V = \int_{-\infty}^{\infty} [\sum_{i=1}^I P(\theta_j)]^2 \phi(\theta_j)d\theta_j - [\int_{-\infty}^{\infty} \sum_{i=1}^I P(\theta_j) \phi(\theta_j)d\theta_j]^2 \quad (7)$$

où $P(\theta_j)$ est la probabilité conditionnelle de répondre correctement à l'item i . Il est à noter que la logique des équations 5 à 7 peut aussi être appliquée aux modèles analysant des items à réponse polychotomique (Culpepper, 2013. p.207).

Évaluer les intégrales présentées aux équations 5 et 7 est un défi important. Voilà pourquoi Dimitrov (2003b) a proposé une méthode numérique pour le cas spécifique du modèle de Rasch. Ici, on peut approximer la variance d'erreur d'un item de cette façon :

$$\sigma_{E_i}^2 \approx A + B \exp \left[-0.5 \left(\frac{b_i}{C} \right)^2 \right] \quad (8)$$

où $A=,011$; $B=,195$; $C=1,797$ quand $|b_i| < 4$ ou $A=,0023$; $B=,171$; $C=2,023$ quand $|b_i| \geq 4$. À l'instar de ce qui est montré à l'équation 6, la somme des $\sigma_{E_i}^2$ devient l'erreur de mesure à un test. De plus, la variance du score vrai pour le nombre de bonnes réponses est calculée de cette façon pour toutes les paires d'items i et k :

$$\sigma_V^2 \approx \sum_{i=1}^I \sum_{k=1}^K \sqrt{[\pi_i(1 - \pi_i) - \sigma_{E_i}^2][\pi_k(1 - \pi_k) - \sigma_{E_k}^2]} \quad (9)$$

en utilisant :

$$\pi_i \text{ ou } k = -0.0114 \frac{1.0228}{1 + \exp(b_i \text{ ou } k / 1.226)}. \quad (10)$$

Notez que Dimitrov (2003a) a développé une stratégie similaire pour calculer la fidélité des scores totaux pour les modèles à réponse dichotomique 1PL, 2PL et 3PL. De leur côté, Raykov, Dimitrov et Asparouhov (2010) ont dérivé une méthode pour calculer les intervalles de confiance. Enfin, Béland et Falk (2022) ont utilisé la quadrature rectangulaire pour estimer les intégrales aux équations 5 et 7 dans le cadre du modèle 2PL. Cette méthode d'estimation a présenté des résultats supérieurs à d'autres coefficients populaires tels que l' α et l'oméga de McDonald (ω) dans le cadre de leur étude de simulation.

4.2. Indice de séparation de personne pour estimer ρ_{Rasch}^2

L'ISP est utilisé dans la majorité des logiciels qui mobilisent le modèle de Rasch (par exemple, RUMM 2030, Winsteps ou la librairie eRm dans R). Ce coefficient calcule l'élément de droite de l'équation 4 de la façon suivante (Wright et Stone, 1999, p. 151) :

$$ISP = 1 - \frac{CMEH}{VH} \tag{11}$$

où $CMEH$ estime σ_E^2 , qui est le carré moyen des erreurs de mesure de l'habileté estimée et VH estime σ_θ^2 , qui est la variance de l'habileté estimée (voir la figure 1, plus bas, pour plus de détails). Le coefficient ISP prend théoriquement une valeur entre zéro et l'unité; et 0,8 est une valeur généralement considérée comme un minimum acceptable (*rule of thumb*) en guise de fidélité (Wright et Masters, 1982, Meyer (2014, p. 90). Il est surtout utile quand le test contient plusieurs items et que l'étendue des habiletés estimées est large (Hojtink et Boomsma, 1995)

Tel que mentionné par Andrich et Marais (2019, p. 130), l'ISP “needs to be interpreted with the distribution of the persons in mind”, alors que le coefficient D_{Rasch} prend sa source dans le paramètre d'items b_i . Enfin, rappelons que Mair *et al.* (2021) soulevaient aussi la distinction entre l'ISP, basé sur les scores logits, et la fidélité des scores totaux: “please note that the concept of reliability and associated problems are fundamentally different between IRT [or Rasch model] and CTT (...). Separation reliability is more like a workaround to make the “change” from CTT to IRT [or Rasch model] easier for users by providing something *familiar*.” (p. 72).

4.3. Synthèse et illustration dans le logiciel R

Les deux stratégies présentées dans cet article sont synthétisées à la figure 1.

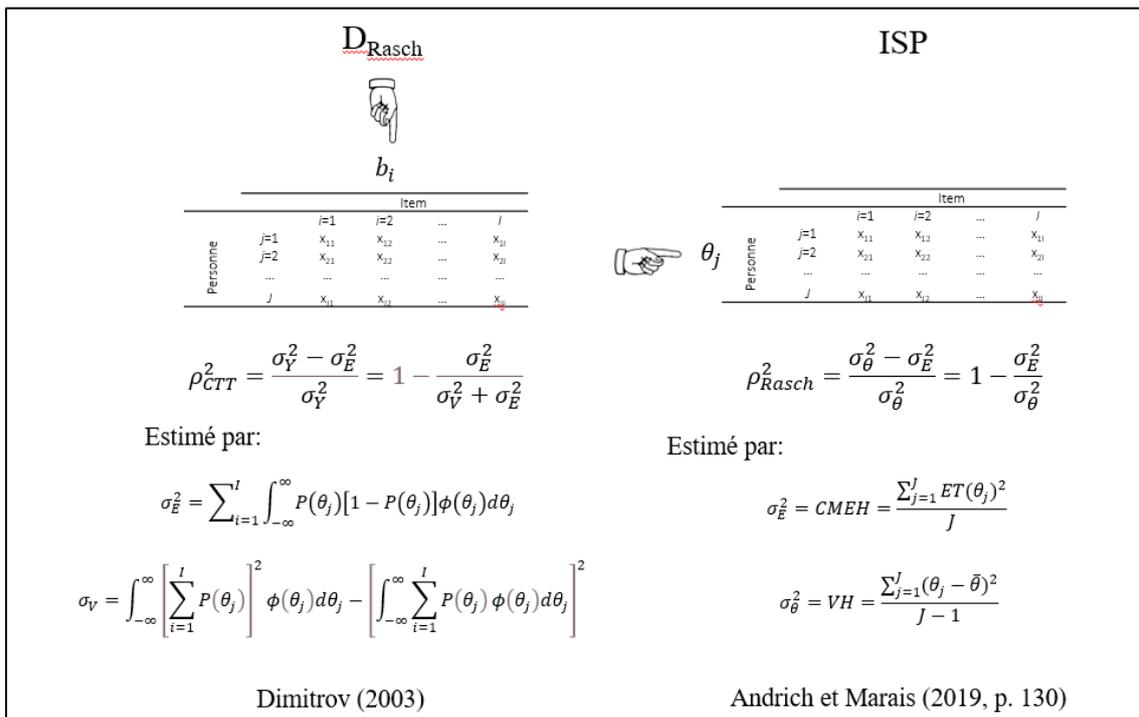


Figure 1 : Méthodes de calcul de D_{Rasch} et l'ISP

Chaque approche s’y prend différemment pour estimer la fidélité: la procédure de Dimitrov (2003b) focalise sur le paramètre d’item b_i , alors que l’ISP calcule la fidélité comme étant conditionnelle aux niveaux mesurés de la variable latente θ_j . Évidemment, la méthode d’estimation des paramètres a un impact sur les valeurs que peuvent prendre ces deux coefficients.

Illustrons, maintenant, leur calcul à partir du logiciel en téléchargement libre R. Premièrement, l’indice D_{Rasch} peut être calculé à l’aide du code présenté aux annexes A et B. En utilisant les valeurs estimées dans le tableau 2 :

Tableau 2 : Valeur estimée de la difficulté

	b_i
Item 1	0,04
Item 2	0,49
Item 3	0,53
Item 4	1,32
Item 5	0,85
Item 6	-0,59
Item 7	-1,26
Item 7	0,15

il suffit d’utiliser la fonction de l’annexe A :

`DARasch(b, Items=8)`

où le vecteur des valeurs du tableau 2, appelé b, permet d’estimer la variance vraie σ_V^2 et la variance d’erreur $\sigma_{E_i}^2$ pour chacun des items :

```
$DRasch
[1] 0.6109597
```

```
$Variances
      [,1]      [,2]      [,3]      [,4]      [,5]
ve 0.20595865 0.19880642 0.19780148 0.15982159 0.18522086
vt 0.04398178 0.04091171 0.04049317 0.02685132 0.03551399
      [,6]      [,7]      [,8]
ve 0.19575264 0.16349044 0.20534247
vt 0.03964954 0.02799373 0.04371092
```

Nous observons que $\sigma_{E_i}^2$ (élément ve dans \$Variances), pour l’item 4, est égal à 0,1598 (voir en caractère gras), soit la variance d’erreur la moins élevée. C’est l’item 1 qui a la σ_V^2 (élément vt) la plus élevée (aussi en caractère gras). Enfin, D_{Rasch} est égal à 0,61, une valeur peu fidèle.

Deuxièmement, l’ISP peut être estimé à l’aide de plusieurs librairies, mais nous allons mobiliser eRm. D’abord, il faut avoir en main des réponses dichotomiques à des items (par exemple en utilisant le code R de l’annexe B). Nous devons estimer les habiletés du modèle avec le code suivant :

```
require(eRm)
pers <- person.parameter(RM(Donnees))
res <- SepRel(pers)
```

où `Donnees` est la matrice de réponses dichotomiques. Ensuite, les deux éléments à estimer dans l'équation 11 peuvent être obtenus. Le carré moyen de l'erreur de mesure de l'habileté estimée s'obtient à l'aide de ce code :

```
CMEH <- sum((pers$se.theta$NAGroup1)^2)/length(pers$se.theta$NAGroup1)
```

alors que la variance des habiletés estimées est obtenue avec ce code:

```
VH <- var(pers$thetapar$NAGroup1)
```

Finalement, l'ISP se calcule de la façon suivante :

```
ISP <- 1 - (CMEH/VH)
```

Dans notre exemple, nous obtenons un $ISP = 1 - (0,79/1,48) = 0,47$, ce qui témoigne d'une fidélité des scores logits faible. Notez que les paramètres étant générés aléatoirement, ce résultat pourrait différer lors de vos propres essais.

5. Une étude de simulation pour illustrer la différence entre DRasch et l'ISP

Pour répondre à notre premier objectif, nous allons maintenant illustrer la différence qui existe entre la méthode D_{Rasch} de Dimitrov (2003a), qui estime la fidélité des scores totaux, et l'ISP, qui estime la fidélité des scores logits. Si ces deux approches de la fidélité sont interchangeable, comme le prétendent certaines personnes avec qui nous déjà avons échangé, leur comportement devrait être le même, n'est-ce pas?

Nous avons généré des matrices de réponses dichotomiques en utilisant le code R inséré dans l'annexe B. Cette approche est décrite dans Béland et Falk (2022) et nous reprenons les paramètres de modèle qui y étaient utilisés pour des raisons de répliquabilité: $b_i \sim N(0, .75^2)$ et $\theta_j \sim N(0,1)$. Ici, deux longueurs de test ont été mobilisées ($I=10, 60$) avec deux tailles d'échantillon ($J=200, 500$). Pour chacune des combinaisons, 1000 ensembles de données dichotomiques ont été générés pour un total de $(2 \times 2) \times 1000 = 4000$ ensembles de données. Pour chacun des ensembles de données, nous avons utilisé la méthode utilisée par Béland et Falk (2022) pour calculer la valeur vraie de la fidélité.

Le coefficient D_{Rasch} a été calculé pour chacun des ensembles de données en utilisant le code R présenté en annexe A. De son côté, le coefficient ISP a été calculé avec la librairie eRm (Mair *et al.*, 2018), qui utilise l'estimateur par maximum de vraisemblance conditionnelle (*conditional maximum likelihood*). Ce qui nous intéresse dans ce court article est de voir à quel point chacun des deux coefficients présente des valeurs qui s'approchent, ou pas, des valeurs réelles.

La figure 2 présente les résultats pour 200 réponses simulées. Nous observons que la corrélation entre les valeurs estimées par D_{Rasch} et les valeurs vraies ($r=0,87$) est fort différente de celle calculée pour l'ISP ($r=0,09$).

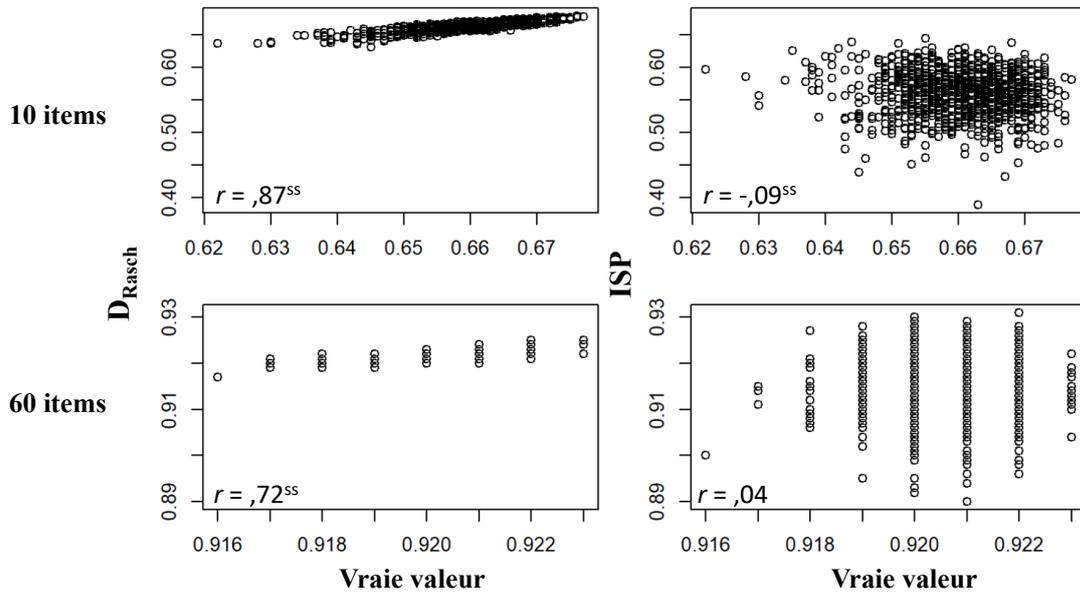


Figure 2: Résultats pour 200 réponses simulées

Les résultats pour 500 patrons de réponses offrent les mêmes tendances que précédemment. Nous observons encore une fois qu'en simulant des réponses à l'aide du modèle de Rasch, les coefficients D_{Rasch} et ISP n'ont pas la même capacité à estimer les valeurs vraies. Par exemple, nous voyons bien qu'avec 10 items, la corrélation entre les scores de D_{Rasch} est de 0,94 alors qu'elle est de -,07 pour l'ISP. En ce qui concerne un test de 60 items, les coefficients de corrélation sont respectivement équivalents à 0,85 et 0,17.

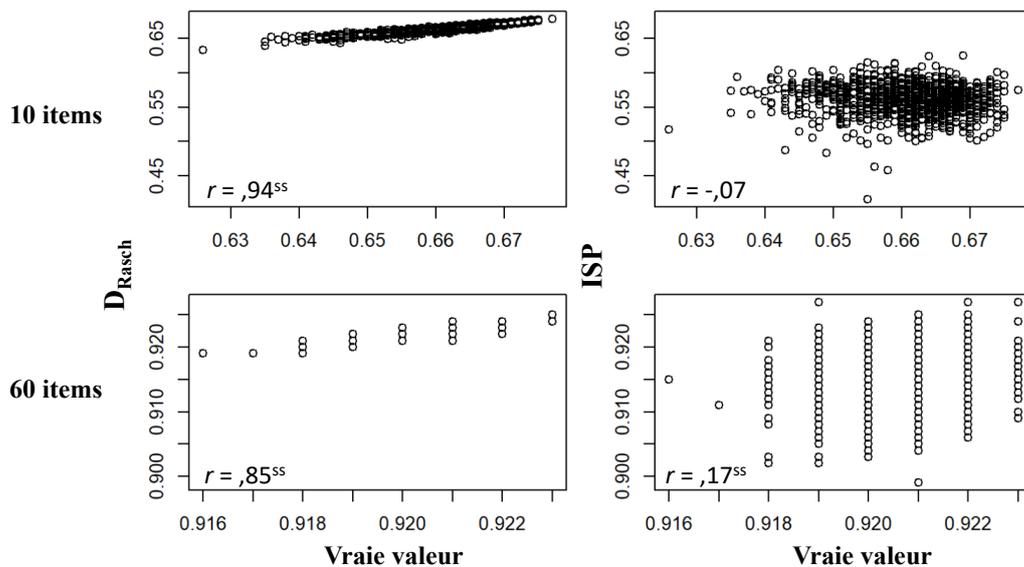


Figure 3: Résultats pour 500 réponses simulées

Ces résultats nous permettent de constater qu'il existe une relation linéaire forte entre D_{Rasch} et les valeurs réelles, alors que ce n'est pas le cas pour l'ISP. Les coefficients sont donc

sensibles à la méthode de simulation utilisée et ne peuvent définitivement par être considérés comme étant interchangeables. Le chercheur doit avoir ce constat en tête au moment de choisir un coefficient ou l'autre.

6. Quoi retenir de tout cela ?

Nous formulons trois constats pour les chercheurs. Premièrement, nous savons depuis longtemps que les coefficients de fidélité ne sont pas des caractéristiques invariables d'un test (Fisher, 1992). Les gens, qui déclarent que certains coefficients de fidélité comme D_{Rasch} et ISP sont interchangeables, devraient revoir leur position à la lumière de la démonstration présentée aux figures 2 et 3.

Deuxièmement, estimer la fidélité est un acte délicat, et il faut argumenter cette propriété psychométrique plutôt que seulement mentionner la valeur d'un coefficient en guise de preuve (Béland et Michelot, 2020; Parkes, 2007). En effet, la fidélité est souvent présentée comme un argument d'autorité générique (pensons à l'utilisation mécanique de l' α de Cronbach dans de très nombreuses études) et il serait pertinent de mettre fin à cette mauvaise façon de faire.

Selon nous, il faut bien argumenter le choix d'utiliser D_{Rasch} et/ou l'ISP dans le cadre d'une étude selon le modèle de Rasch. Et cela veut dire que le contexte du testing est un élément fondamental de l'argumentaire. Par exemple, si un chercheur souhaite comparer les résultats d'une analyse à l'aide du modèle de Rasch à d'autres études où uniquement le coefficient α de Cronbach a été utilisé, il sera préférable qu'il utilise D_{Rasch} , à l'ISP car ce coefficient est aussi basé sur la fidélité des scores totaux. Si le chercheur veut plutôt produire plutôt une méta-analyse intégrant des informations sur la fidélité du modèle de Rasch, il trouvera beaucoup plus de références à l'ISP dans les écrits scientifiques et ce coefficient sera plus intéressant.

Troisièmement, dans l'optique de discuter des propriétés éduométriques des items, le coefficient D_{Rasch} présente l'avantage important d'être plus près de la perspective la plus connue : la fidélité des scores totaux. Celui-ci permet d'obtenir des valeurs $\sigma_{V_i}^2$ et $\sigma_{E_i}^2$ pour chacun des items, alors que l'ISP tire sa source dans l'habileté estimée dans la version retenue dans cet article. D_{Rasch} est donc conceptuellement plus cohérent avec la définition ρ_{TCT}^2 , qui nous semble aussi plus facile à comprendre pour les non-initiés.

Enfin, nous croyons que ces coefficients doivent être étudiés de plus près (par exemple avec des données réelles) et il est maintenant temps de réengager les chercheurs sur la question de la fidélité dans le cadre d'analyses mobilisant le modèle de Rasch.

7. Bibliographie

- Andrich, D. et Marais, I. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Springer Nature Singapore Pte Ltd.
- Baker, F.B. et Kim, S.H. (2004). *Item Response Theory Parameter Estimation Techniques*. 2nd Edition, CRC Press, Boca Raton.
- Bartholomew, D. J., Bassin, E. L. et Schuessler, K. F. (1993). Properties of a Latent Trait Reliability Coefficient. *Sociological Methods & Research*, 22(2), 163-192. <https://doi.org/10.1177/0049124193022002001>
- Béland, S. et Falk, C. F. (2022). A Comparison of Modern and Popular Approaches to Calculating Reliability for Dichotomously Scored Items. *Applied Psychological Measurement*, 46(4), 321-337. <https://doi.org/10.1177/01466216221084210>
- Béland, S. et Michelot, F. (2020). Une note sur le coefficient omega (ω) et ses déclinaisons pour estimer la fidélité des scores. *Mesure et évaluation en éducation*, 43(3), 103–122: <https://doi.org/10.7202/1084526ar>
- Culpepper, S.A. (2013) The Reliability and Precision of Total Scores and IRT Estimates as a Function of Polytomous IRT Parameters and Latent Trait Distribution. *Applied Psychological Measurement*, 37(3), 201-225. <https://doi.org/10.1177/0146621612470210>
- Dionne, É. et Béland, S. (Éds, 2023). *Appliquer le modèle de Rasch : problèmes et pistes de solution*. Presses de l'Université du Québec.
- Dimitrov, D. M. (2003a). Marginal True-Score Measures and Reliability for Binary Items as a Function of Their IRT Parameters. *Applied Psychological Measurement*, 27(6), 440–458. <https://doi.org/10.1177/0146621603258786>
- Dimitrov, D. M. (2003b). Reliability and True-Score Measures as a Function of Their Rasch difficulty Parameter. *Journal of applied measurement*, 4(3), 222-233.
- Dimitrov, D. (2023). *D-scoring Method of Measurement: Classical and Latent Frameworks* (1st ed.). Routledge. <https://doi.org/10.4324/9781003343004>
- Finkel, E. J., Eastwick, P. W. et Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology*, 108(2), 275-297. <https://doi.org/10.1037/pspi0000007>
- Fisher, W. Jr. (1992). Reliability, Separation, Strata Statistics. *Rasch Measurement Transactions*, 6, 238.
- Harvill, L. M. (1991). NCME Instructional Module: Standard Error of Measurement. *Educational measurement, issues and practice*, 10, 33-41.
- Hoijtink, H. et Boomsma, A. (1995). On person parameter estimation in the dichotomous Rasch model. In GH. Fischer, et IW. Molenaar (Eds.), *Rasch models: foundation, recent developments and applications*. (p. 53-68). Springer.
- Lord, F.M. et Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Mair, P., Hatzinger, R. et Maier M. J. (2021). *eRm: Extended Rasch Modeling*. 0.16-1. <http://erm.r-forge.r-project.org/>
- Martineau, G. (1982). Exploration des valeurs possibles du coefficient α de Cronbach. *Revue des sciences de l'éducation*, 8(1), 135-143. <https://doi.org/10.7202/900362ar>
- Meyer, J. P. (2014). *Applied measurement with jMetrik*. New York, Routledge.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Institute of Educational Research.

- Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26, 2-10.
- Raykov, T., Dimitrov, D., D. et Asparouhov, T. (2010) Evaluation of Scale Reliability With Binary Measures Using Latent Variable Modeling. *Structural Equation Modeling*, 17(2), 265-279. <https://doi.org/10.1080/10705511003659417>
- Sijtsma K. (2016). Playing with Data - Or How to Discourage Questionable Research Practices and Stimulate Researchers to Do Things Right. *Psychometrika*, 81, 1-15. <https://doi.org/10.1007/s11336-015-9446-0>
- Wright B. D. et Masters G. N. (1982). *Rating scale analysis*. Chicago, Illinois: MESA Press.
- Wright, B. D. et Stone, M. H. (1999). *Measurement Essentials*. Wide Range, Inc.

8. Annexes

8.1. Annexe A

```
# Fonction pour calculer la méthode de Dimitrov (2003b)

DRasch <- function(b,Items) {

  out<-matrix(NA,4,Items,byrow = TRUE)
  rownames(out) <- c("b","p","ve","vt")

  for(i in 1:Items) {

# Item

  out[1,i] <- b[i]
  out[2,i] <- -.0114+(1.0228/(1+exp(out[1,i]/1.226))) # p

  if (out[1,i] < abs(4)) {out[3,i] <- .0110+.195*exp(-.5*((out[1,i]/1.797)^2))} # Ve
  else {out[3,i] <- .0023+.171*exp(-.5*((out[1,i]/2.023)^2))} # Ve

  out[4,i] <- out[2,i]*(1-out[2,i])-out[3,i] # Vt
  if (out[4,i] < 0) {out[4,i] = 0}

# Test

  veT <- sum(out[3,])
  vtT <- sum(sqrt(as.numeric(out[4,])%*%t(as.numeric(out[4,])))
  RealTest <- vtT / (vtT+veT)

  RRel <- vtT / (vtT+veT)

}

  output <- list()
  output$DRasch<-RRel
  output$Variances<-out[3:4,]
  return(output)

}
```

8.2. Annexe B

```
# Fonction pour générer des réponses dichotomiques

dataRasch<-function(nSubjs,nItems,th,b){

  pr<-function(th,b) exp(th-b)/(1+exp(th-b))

  res<-matrix(NA,(nSubjs),nItems)

  for (i in 1:nSubjs) res[i,]<-rbinom(nItems,1,pr(th[i],b))

  return(res)

}

nSubjs <- 200 # 200 répondants
nItems <- 8 # 8 items
th <- rnorm(nSubjs) # Paramètre  $\theta_j$  de l'équation 3
b <- rnorm(nItems) # Paramètre  $b_i$  de l'équation 3, utilisé dans la fonction DRasch

# Enfin, la génération de la matrice de réponses dichotomiques :

Donnees <- dataRasch(nSubjs,nItems,th,b)
```

8.3. Annexe C

```
require(eRm)
pers <- person.parameter(RM(Donnees))
res <- SepRel(pers)
```