Received: 12 February 2025 | Accepted: 9 June 2025

DOI: 10.1111/bmsp.70002

ARTICLE

the british psychological society
promoting excellence in psychology

Inferences of associated latent variables by the observable test scores

Rudy Ligtvoet

University Hospital Cologne, Cologne, Germany

Correspondence
Rudy Ligtvoet, Klinik I für Innere Medizin, Uniklinik Köln, 50924 Köln, Germany.
Email: rudy.ligtvoet@uk-koeln.de

Abstract
Test scores, like the sum score, can be useful for making inferences about the latent variables. The conditions under which such test scores allow for inferences of the latent variables based on a "weaker" stochastic ordering are generalized to any monotone latent variable model for which the latent variables are associated. The generality of these conditions places the sum score, or indeed any test score, well beyond a mere intuitive measure or a relic from classical test theory.

KEYWORDS
monotone likelihood ratio, sum score, variables being associated

1 | INTRODUCTION

Latent variable models for multiple item scores are useful to the extent that they allow for inferences to be made about the latent variables $\Theta_1, \ldots, \Theta_d$, that are assumed to describe the dependencies that exist between the item scores. Many parametric latent variable models are available that allow for such inferences. A downside of these models is their reliance on parametric assumptions that are usually made based on pragmatic considerations of mathematical convenience. Hereto, a notable exception is the derivation of the Rasch (1960) model from the principle of *specific objectivity* (Rasch, 1977), for which the parametric model characteristics follow from requiring the sum score across items to be a sufficient statistic for the *unidimensional* latent variable (i.e., $d = 1$; Andersen, 1977; Fischer, 1974). Results of testing are in practice also communicated in terms of the same sum score, but often without the empirical validation of any particular latent variable model. Unlike the parametric model restrictions, this presents a pragmatic consideration of a different kind, namely the convenience with which the sum score on a test can be computed and communicated. Such use of the sum score is often referred to as measurement by *fiat* (Torgerson, 1958) and a relic from a classical treatment of test scores (e.g., McNeish & Wolf, 2020; Sijtsma et al., 2024; Widaman & Revelle, 2023, for a discussion). However, the assessment of the sum score as a mere intuitive measure ignores the bulk of evidence that has been accumulated in favour of the use of the sum score for making inferences in the context of latent variable models. For example, Ellis and Junker (1997) provided a fully observable characterization of a general class of latent variable models for an infinite sequence of random item variables $(X_1, X_2, \ldots)$, whereby the latent variables can be consistently estimated from any sequence with an arbitrary number of random item variables omitted. This work is closely related to the concept of *essential unidimensionality* of the sequence $(X_1, X_2, \ldots)$, where a dominant latent variable can be consistently estimated by the sum score (Junker, 1991; Stout, 1990). For a finite number of n random item variables $X_1, \ldots, X_n$, the focus has been on attaining a *stochastic ordering* of the unidimensional latent variable Θ by the sum score $S = X_1 + \cdots + X_n$. This property is referred to as SOL (Hemker et al., 1997) and states that, for all θ,

$$P[\Theta > \theta | S = s] \text{ is non-decreasing in } s.$$

As SOL is hard to establish directly, a more restrictive property called a *monotone likelihood ratio* (MLR) is usually considered instead. The *congeneric one-factor model* (Jöreskog, 1971) is an example of a model for normal item scores that implies this MLR (Ligtvoet, 2022a), and thus allows for a stochastic ordering on the latent variable Θ by the sum score S. Mokken's (1971) model of *monotone homogeneity* for binary item variables, and special cases thereof, like the *two-parameter logistic model* (Birnbaum, 1968), the *normal ogive models* (Lord, 1952, 1980), and the Rasch (1960) model, all imply an MLR (Grayson, 1988; Ünlü, 2008). However, Hemker et al. (1996, 1997) found that most of the latent variable models that are used in practice for polytomous item scores do not imply an MLR, which include the *graded response model* (Samejima, 1969), the *generalized partial credit model* (Muraki, 1992), and the *sequential model* (Tutz, 1990). This suggests that the restrictions imposed by the MLR property may be too restrictive. To mitigate the restrictive nature of an MLR, Van der Ark and Bergsma (2010) proposed a "weaker" version of SOL, and showed that it is implied by all the above mentioned latent variable models.

In the following, two results are discussed that support the use of the test score for making stochastic inferences. The first result is by Ghurye and Wallace (1959), who provide sufficient conditions for an MLR. The result by Jogdeo (1978) provides the conditions under which any test score that is a non-decreasing function of the item variables, including the sum score, implies a weak SOL for any non-decreasing function of the latent variables $\Theta_1, \ldots, \Theta_d$. This second result considerably generalizing the class of latent variable models that allow for a stochastic ordering by the test score. More specifically, Theorem 2 presents the general (non-parametric) conditions under which the observable test score is associated with any and each of the multi-dimensional latent variables, thus placing the sum score well beyond a mere intuitive measure or a relic from classical test theory.

2 | STOCHASTIC ORDERINGS ON LATENT VARIABLES

The first result on the stochastic ordering of the unidimensional latent variable Θ by the sum score S, in accordance to an MLR, was obtained by Ghurye and Wallace (1959).

2.1 | A monotone likelihood ratio

Let f be a real-valued positive function defined on $\mathcal{X} \times \mathbb{R}$, where $\mathcal{X}$ is an ordered set and $f(u, \theta)$ is taken to be measurable in u for each θ. We say that the function f is (latent) *totally positive of order 2* (TP; Holland & Rosenbaum, 1986; Karlin, 1968), if for all $u_1, u_2 \in \mathcal{X}$ with $u_1 \le u_2$, and all $\theta_1 \le \theta_2$,

$$f(u_1, \theta_1) f(u_2, \theta_2) \ge f(u_1, \theta_2) f(u_2, \theta_1).$$

Further, we say that f is a *Pólya frequency function of order 2* (PF; e.g., Efron, 1965), if for all $u_1, u_2, u_3, u_4 \in \mathcal{X}$ with $u_1 \le u_2$ and $u_3 \le u_4$, and all θ,

$$f(u_1 - u_3, \theta) f(u_2 - u_4, \theta) \ge f(u_1 - u_4, \theta) f(u_2 - u_3, \theta).$$

Let the convolution $f * g$ be defined as

$$(f * g)(u, \theta) = \int f(u - v, \theta) g(v, \theta) \; dv.$$

Here, the importance of the convolution stems from the fact that the density (or mass function) of the sum $U + V$ of two independent random variables U and V is described by the convolution of their densities.

Theorem 1. The convolution $f * g$ is both TP and PF, if each of the functions f and g are both TP and PF.

(Ghurye & Wallace, 1959)

Remark. To appreciate the scope of Theorem 1, consider U and V to be the real-valued random variables with densities $f(u, \theta)$ and $g(v, \theta)$, that are independent, conditional on Θ. That is, for all values u and v, and all θ,

$$P[U \leq u, V \leq v | \Theta = \theta] = P[U \leq u | \Theta = \theta] P[V \leq v | \Theta = \theta].$$

If each of the functions f and g are both TP and PF, then according to Theorem 1, so is the density $(f * g)(u + v, \theta)$. Next, consider the random variables $X_1, \ldots, X_n$ that are independent, conditional on Θ, and have densities that are both TP and PF. Then, sequentially taking $U = X_1 + \cdots + X_{i-1}$ and $V = X_i$, for $i = 2, \ldots, n$, yields that $f(s, \theta)$ is TP (and PF), for $S = X_1 + \ldots + X_n$. Finally, if $f(s, \theta)$ is TP, then this is said to correspond to an MLR of Θ by S.

The reason many of the polytomous latent variable models do not imply an MLR is that they do not satisfy the PF requirement (Ligtvoet, 2012). The *rating scale model* (Andrich, 1978) is an example of a model that satisfies both TP and PF, and therefor implies an MLR. However, Masters' (1982) partial credit model does not imply the PF property, yet it implies an MLR, as it has the sum score as a sufficient statistic for the latent variable. This shows that the conditions in Theorem 1 are not necessary for an MLR, albeit sufficient.

2.2 | Associated random variables

The weak SOL property proposed by Van der Ark and Bergsma (2010) is less restrictive than an MLR and corresponds to the property of two random variables being *positive quadrant dependent* (PQD; Lehmann, 1966). The two random variables U and V are PDQ, if for all values u and v

$$P[U \leq u, V \leq v] \geq P[U \leq u] \; P[V \leq v], \tag{1}$$

from which the weak SOL property is obtained by taking $U = \Theta$ and $V = X_1 + \ldots + X_n$.

The second result describes a stochastic ordering in accordance with PQD of any two non-decreasing functions g and h. Here, the function g pertains to the *associated* (multi-dimensional) random vector $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_d)$. The function $g: \mathbb{R}^d \rightarrow \mathbb{R}$ is said to be non-decreasing, whenever $g(\boldsymbol{\theta})$ is non-decreasing in each element $\theta_1, \ldots, \theta_d$, for all $\boldsymbol{\theta}$. The function $h(\boldsymbol{X})$ defined on $\boldsymbol{X} = (X_1, \ldots, X_n)$ can be interpreted as any test score that is non-decreasing in the item scores, which includes but is not restricted to the sum score. For example, $h(\boldsymbol{X})$ can denote any discretization of the sum score or the positively weighted average $(a_1 X_1 + \cdots + a_n X_n)/n$, for fixed $a_i \geq 0$ (Ligtvoet, 2022a; Rosenbaum, 1984).

First, to describe the dependencies between the random item variables, we say that $X_1, \ldots, X_n$ are *conditionally independent* (CI), given $\boldsymbol{\Theta}$, if

$$P[X_1 \leq x_1, \ldots, X_n \leq x_n | \boldsymbol{\Theta} = \boldsymbol{\theta}] = P[X_1 \leq x_1 | \boldsymbol{\Theta} = \boldsymbol{\theta}] \cdots P[X_n \leq x_n | \boldsymbol{\Theta} = \boldsymbol{\theta}],$$

for all $x_1, \ldots, x_n$ and $\boldsymbol{\theta}$. Also, we say that the assumption of *monotonicity* (M) is satisfied, if

$$P[X_i > x_i | \boldsymbol{\Theta} = \boldsymbol{\theta}] \text{ are non-decreasing in } \boldsymbol{\theta},$$

for all x_i and $i = 1, \ldots, n$. Assumption M relaxes the TP requirement for an MLR (Holland & Rosenbaum, 1986). Second, we say that the random vector $\boldsymbol{U}$ is associated (Esary et al., 1967), if

$$\text{Cov}[g(\boldsymbol{U}), h(\boldsymbol{U})] \geq 0,$$

for any non-decreasing functions g, h. The following result was proven by Jogdeo (1978).

Theorem 2. The random vector $\boldsymbol{U} = (\boldsymbol{X}, \boldsymbol{\Theta})$ is associated, whenever the following three conditions hold*:*

(Jogdeo, 1978)

a. $\boldsymbol{\Theta}$ *is associated,*
b. $\boldsymbol{X}$ *is conditionally associated given* $\boldsymbol{\Theta}$, *and*
c. $E[h(\boldsymbol{X}) | \boldsymbol{\Theta} = \boldsymbol{\theta}]$ *is non-decreasing in* $\boldsymbol{\theta}$, *for any non-decreasing function* h.

Remark. The condition (b) is implied by CI (Esary et al., 1967, theorem 2.1), and CI and M together imply condition (c) (Holland & Rosenbaum, 1986, lemma 2). Hence, CI and M, together with (a) imply that $(\boldsymbol{X}, \boldsymbol{\Theta})$ is associated. Also, as any subset of associated random variables is associated, CI, M, and $\boldsymbol{\Theta}$ being associated, imply that $\boldsymbol{X}$ is also associated (Holland & Rosenbaum, 1986, theorem 8). Finally, $(\boldsymbol{X}, \boldsymbol{\Theta})$ associated implies that $\text{Cov}[g(\boldsymbol{\Theta}), h(\boldsymbol{X})] \geq 0$, for all g, h non-decreasing, which in turn implies PQD, as can be obtain from Esary et al. (1967, theorem 4.4), by taking $U = g(\boldsymbol{\Theta})$ and $V = h(\boldsymbol{X})$.

The above shows that, for any latent variable model for which CI and M are satisfied, and $\boldsymbol{\Theta}$ is associated, PQD holds for any $U = g(\boldsymbol{\Theta})$ and $V = h(\boldsymbol{X})$ in (1), with g, h non-decreasing. This mean that for any latent variable model for which these conditions are satisfied, any test score $h(\boldsymbol{X})$ that is a non-decreasing function of the item variables (including the sum score) provides a weak stochastic ordering of any non-decreasing function of the latent variables.

3 | DISCUSSION

Theorem 2 generalizes the condition that allow for stochastic inferences to be made about latent variables to any non-decreasing function of the observable random variables $X_1, \ldots, X_n$. These conditions for a PQD are not confined to unidimensional latent variable models, but apply to multiple latent variables that are associated. Examples of such models include the multiple factor analysis model and hierarchical factor model, with non-negative factor loadings and non-negative correlations between the latent factor (Ellis, 2015; Krijnen, 2004). The assumption that the latent variables $\Theta_1, \ldots, \Theta_d$ are associated, together with M and CI, means that any non-decreasing function of these latent variables is positively related to any test score $h(\boldsymbol{X})$ in terms of a PQD. In turn, the same test score allows for stochastic inferences in terms of a PQD about any one of the latent variables $\Theta_1, \ldots, \Theta_d$. So, under the conditions in Theorem 2, the test score does not reveal anything about the latent structure other than that any non-decreasing function of these latent variables has a covariance with the test score that is non-negative; a minimal requirement of a test score to be considered useful. Additional (parametric) assumptions would therefore be required to make specific statements about the structure of the latent variables that account for the dependencies between the item score variables (e.g., Ellis et al., 2025; Ellis & Sijtsma, 2023). But whichever additional restrictions are imposed on the latent structure, these should agree with the conditions in Theorem 2 in order for the test score to be considered useful for making inferences about the ordering on the latent variables.

Although PQD is a weaker version of SOL, the generality of the conditions under which it allows for inferences of the latent variables from the observed item scores places the sum score, or indeed any test score, well beyond a mere intuitive measure or a relic from classical test theory. In their discussion on the use of the sum score, McNeish and Wolf (2020) mention that their goal is *"to raise awareness that sum scoring requires rather strict constraints, imposing these constraints requires the same type of justification as any other latent variable model, and sum scoring corresponds to a statistical model and is not a model-free arithmetic calculation"* (p. 2287). The results of Theorem 2 largely agree with this assessment, but show that the conditions that allow for the use of the sum score for making inferences are far more general that those considered in their paper. The conditions for a PQD also imply the observable property that $\boldsymbol{X}$ is associated. A test for the item scores being associated could provide an empirical justification for the use of the sum score. Unfortunately, this property is hard to directly ascertain in practice (Ligtvoet, 2022b; Walkup, 1968) and in need of further investigation (Ligtvoet, 2023).

AUTHOR CONTRIBUTIONS

Rudy Ligtvoet: writing – original draft; writing – review and editing; validation; methodology; conceptualization; investigation; formal analysis.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

DISCLOSURE OF ARTIFICIAL INTELLIGENCE-GENERATED CONTENT (AIGC) TOOLS

No AIGC tools were used.

ORCID

Rudy Ligtvoet https://orcid.org/0000-0001-6546-0911

REFERENCES

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*(1), 69–81. https://doi.org/10.1007/BF02293746

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. https://doi.org/10.1007/BF02293814

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick *Statistical theories of mental test scores* (pp. 396–479). Addison-Wesley.

Efron, B. (1965). Increasing properties of Pólya frequency function. *The Annals of Mathematical Statistics*, *36*(1), 272–279. https://www.jstor.org/stable/2238092

Ellis, J. L. (2015). MTP2 and partial correlations in monotone higher-order factor models. In R. E. Millsap, D. M. Bolt, L. A. Van der Ark, & W. C. Wang (Eds.), *Quantitative psychology research* (pp. 261–272). Springer. https://doi.org/10.1007/978-3-319-07503-7_16

Ellis, J. L., & Junker, B. W. (1997). Tail-measurability in monotone latent variable models. *Psychometrika*, *62*(4), 495–523. https://doi.org/10.1007/BF02294640

Ellis, J. L., & Sijtsma, K. (2023). A test to distinguish monotone homogeneity from monotone multifactor models. *Psychometrika*, *88*(2), 387–412. https://doi.org/10.1007/s11336-023-09905-w

Ellis, J. L., Van der Ark, L. A., & Sijtsma, K. (2025). An overall test of pairwise mean conditional covariances in IRT. *Psychometrika*, *90*(1), 384–414. https://doi.org/10.1017/psy.2024.21

Esary, J. D., Proschan, F., & Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics*, *38*(5), 1466–1474. https://doi.org/10.1214/aoms/1177698701

Fischer, G. H. (1974). *Einführung in die theorie psychologischer tests*. Verlag Hans Huber.

Ghurye, S. G., & Wallace, D. L. (1959). A convolutive class of monotone likelihood ratio families. *The Annals of Mathematical Statistics*, *30*(4), 1158–1164. https://doi.org/10.1214/aoms/1177706101

Grayson, D. A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika*, *53*(3), 383–392. https://doi.org/10.1007/BF02294219

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, *61*(4), 679–693. https://doi.org/10.1007/BF02294042

Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, *62*(3), 331–347. https://doi.org/10.1007/BF02294555

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, *14*(4), 1523–1543. https://doi.org/10.1214/aos/1176350174

Jogdeo, K. (1978). On a probability bound of Marshall and Olkin. *The Annals of Statistics*, *6*(1), 232–234. https://doi.org/10.1214/aos/1176344082

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. https://doi.org/10.1007/BF02291393

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika*, *56*(2), 255–278. https://doi.org/10.1007/BF02294462

Karlin, S. (1968). *Total positivity*. Stanford University Press.

Krijnen, W. P. (2004). Positive loadings and factor correlations from positive covariance matrices. *Psychometrika*, *69*(4), 655–660. https://doi.org/10.1007/BF02289861

Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, *37*(5), 1137–1153. https://doi.org/10.1214/aoms/1177699260

Ligtvoet, R. (2012). An isotonic partial credit model for ordering subjects on the basis of their sum scores. *Psychometrika*, *77*(3), 479–494. https://doi.org/10.1007/s11336-012-9272-6

Ligtvoet, R. (2022a). The sum scores and discretization of variables under the linear normal one-factor model. In M. Wiberg, D. Molenaar, J. González, J.-S. Kim, & H. Hwang (Eds.), *Quantitative psychology. IMPS 2021. Springer proceedings in mathematics & statistics* (Vol. 393, pp. 227–235). Springer. https://doi.org/10.1007/978-3-031-04572-1_17

Ligtvoet, R. (2022b). Incomplete tests of conditional association for the assessment of model assumptions. *Psychometrika*, *87*(4), 1214–1237. https://doi.org/10.1007/s11336-022-09841-1

Ligtvoet, R. (2023). A Bayesian test for the association of binary response distributions. In L. A. Van der Ark, W. H. M. Emons, & R. R. Meijer (Eds.), *Essays on contemporary psychometrics. methodology of educational measurement and assessment*. Springer. https://doi.org/10.1007/978-3-031-10370-4_14

Lord, F. M. (1952). *A theory of test scores*. Psychometric monograph no. 7. Psychometric Society.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*, 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. https://doi.org/10.1177/014662169201600206

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Rasch, G. (1977). On specific objectivity an attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, *14*(1), 58–94.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika*, *49*(3), 425–435. https://doi.org/10.1007/BF02306030

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(Suppl 1), 1–97. https://doi.org/10.1007/BF03372160

Sijtsma, K., Ellis, J. L., & Borsboom, D. (2024). Recognize the value of the sum score, Psychometrics' greatest accomplishment. *Psychometrika*, *89*(1), 84–117. https://doi.org/10.1007/s11336-024-09964-7

Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*(2), 293–325. https://doi.org/10.1007/BF02295289

Torgerson, W. S. (1958). *Theory and methods of scaling*. Wiley.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, *43*(1), 39–55. https://doi.org/10.1111/j.2044-8317.1990.tb00925.x

Ünlü, A. (2008). A note on monotone likelihood ratio of the total score variable in unidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 179–187. https://doi.org/10.1348/000711007X173391

Van der Ark, L. A., & Bergsma, W. P. (2010). A note on stochastic ordering of the latent trait using the sum of polytomous item scores. *Psychometrika*, *75*(2), 272–279. https://doi.org/10.1007/s11336-010-9147-7

Walkup, D. W. (1968). Minimal conditions for association of binary variables. *SIAM Journal on Applied Mathematics*, *16*(6), 1394–1403. https://doi.org/10.1137/0116115

Widaman, K. F., & Revelle, W. (2023). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788–806. https://doi.org/10.3758/s13428-022-01849-w

How to cite this article: Ligtvoet, R. (2026). Inferences of associated latent variables by the observable test scores. *British Journal of Mathematical and Statistical Psychology*, *79*, 139–145. https://doi.org/10.1111/bmsp.70002