A Method for Investigating the Intersection of Item Response Functions in Mokken's Nonparametric IRT Model. — Source link ↗

Klaas Sijtsma, Rob R. Meijer

Institutions: VU University Amsterdam

Related papers:

- A Theory and Procedure of Scale Analysis

- Methodology Review : Evaluating Person Fit

- Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics

- A nonparametric approach to the analysis of dichotomous item responses

- A survey of theory and methods of invariant item ordering

Share this paper: 😊 🐦 in ✉

View more about this paper here: https://typeset.io/papers/a-method-for-investigating-the-intersection-of-item-response-4hf9idlchp

A Method for Investigating the Intersection of Item Response Functions in Mokken's Nonparametric IRT Model

Klaas Sijtsma and Rob R. Meijer
Vrije Universiteit

For a set of k items having nonintersecting item response functions (IRFs), the H coefficient (Loevinger, 1948; Mokken, 1971) applied to a transposed persons by items binary matrix H^T has a non-negative value. Based on this result, a method is proposed for using H^T to investigate whether a set of IRFs intersect. Results from a monte carlo study support the proposed use of H^T. These results support the use of H^T as an extension to Mokken's nonparametric item response theory approach. *Index terms: double monotonicity, H^T coefficient, intersection of item response functions, item response theory, Mokken models, nonparametric models.*

In item response models, the item response function (IRF) specifies the relationship between the probability of a correct or keyed response on a dichotomous item and the measurement scale of the psychological attribute under consideration. Usually, this probability is a nondecreasing function of the scale. Mokken (1971; Mokken & Lewis, 1982) proposed two item response models for ordering n persons and k dichotomous items on a measurement scale. Because neither IRFs nor distributions of parameters are parametrically defined, both models can be characterized as nonparametric (e.g., Meijer, Sijtsma, & Smid, 1990).

The Mokken (1971) model of monotone homogeneity (MH) is based on three assumptions: unidimensional measurement, local stochastic independence of responses to items, and nondecreasing IRFs. Similar models are discussed by Holland (1981), Rosenbaum (1984, 1987a), Holland and Rosenbaum (1986), Stout (1987), and Grayson (1988). The MH model is relevant because of its potential for ordering persons along the measurement scale, denoted by θ. Because IRFs are not parametrically defined, θ cannot be estimated numerically (see Lewis, 1983, and Mokken & Lewis, 1982, for the estimation of a modified person parameter). Because the orderings based on the number-correct true score T from classical test theory and θ are identical (Lord, 1980; Mokken, 1971, p. 128), T may replace θ for ordering persons.

The Mokken (1971) model of double monotonicity (DM) is based on the assumptions of the MH model as well as the assumption that the IRFs do not intersect. The shape of the IRFs is free to vary within these limits, but IRFs for larger numbers of items are likely to be more parallel than for smaller numbers. The assumption of nonintersecting IRFs implies that for each measurement value θ, the ordering of the success probabilities on k items is the same (see Rosenbaum, 1987b). Ties occur if IRFs touch. An invariant ordering of item difficulties is important, for example, in establishing rules for beginning and ending subtest sessions in intelligence testing of children (e.g., Bleichrodt, Drenth, Zaal, & Resing, 1985). An application of the property of invariant item ordering in different groups is found in item bias research (Hulin, Drasgow, & Parsons, 1983, p. 162). Finally, an invariant item

ordering is required for the application of a reliability method proposed by Mokken (1971; Sijtsma & Molenaar, 1987).

In logistic item response theory (IRT) models, the item difficulty δ is defined as that value of θ for which the probability of success lies exactly halfway between the lower and upper asymptotes of the IRF (Hambleton & Swaminathan, 1985, p. 38; Lord, 1980, pp. 12–13). This definition of δ is not suitable for models that do not have logistic IRFs, such as the DM model. For example, for two items g and h, let the IRF of g have larger values for all θs. Obviously, item g would be considered to be the easier item. However, if the minimum success probabilities of g and h were unequal (e.g., .5 and 0.0, respectively), then within the DM model it is possible to conceive of IRFs such that using the definition of δ for logistic IRFs $\delta_g > \delta_h$.

Alternatively, with the DM model δ can be defined as the value of θ for which the success probability is the same constant for all items (e.g., Mokken, 1971, p. 174). The property of nonintersecting IRFs then can be stated as follows: If for two items g and h, $\delta_g > \delta_h$, then for all θs the conditional probability of a correct response on h is at least as large as on g (Meijer et al., 1990; Mokken, 1971, p. 118). Because the IRFs are not parametrically defined, δ cannot be estimated numerically. It can be shown (Mokken, 1971, p. 126), however, that δ and the proportion of correct responses on an item g (denoted by π_g) are reversely ordered. Because π_g can be estimated consistently and without bias from empirical data, it can replace δ to estimate the ordering of items according to their difficulties.

Mokken (1971) used Loevinger's (1948) scalability coefficient H as an indicator of measurement quality for use with the MH model. Specifically, Mokken (1971, p. 150) demonstrated that for k items having nondecreasing IRFs,

$$0 \leq H \leq 1 \quad , \tag{1}$$

where $H = 0$ if and only if at least $k - 1$ IRFs are constant functions of θ (a trivial case of unsuccessful measurement). A non-negative value of H constitutes a necessary condition for the nondecreasingness of IRFs: the assumption is rejected with a negative value of H.

Despite some critical comments on the role of the H coefficient in Mokken's models (e.g., Roskam, van den Wollenberg, & Jansen, 1986; see Mokken, Lewis, & Sijtsma, 1986, for a critical reply), the H coefficient has proven to be a useful tool for test construction. Examples are provided by Mokken and Lewis (1982) for quality of sleep, Kingma and TenVergert (1985) for Piagetian conservation tasks, and Meijer et al. (1990) for verbal analogies.

The previous research using H was concerned with the nondecreasingness of IRFs. In the present study, the H coefficient based on the transposed data matrix H^T was used to investigate whether IRFs intersect. In particular, it is shown that for k items having nonintersecting IRFs,

$$0 \leq H^T \leq 1 \quad , \tag{2}$$

where $H^T = 0$ if and only if all IRFs coincide, except for at most one value of θ. Analogously to Equation 1, non-negative values of H^T constitute a necessary condition for nonintersection of IRFs and, therefore, negative values are in conflict with this assumption.

The H^T coefficient fills a gap that exists in the study of whether two or more IRFs intersect. Methods proposed by Mokken (1971, pp. 180–182) and Rosenbaum (1987b) are laborious to apply, because they involve the visual inspection of many orderings. Another deficiency of these methods is that the information obtained about intersections is often difficult to combine into one uniform decision about model fit. The application of H^T proposed here reduces these problems.

Transposition of the Data Matrix: The H^T Coefficient

Items are indexed by g and h ($g, h = 1, \ldots, k$), and persons are indexed by i and j ($i, j = 1, \ldots, n$). Let π_g denote the proportion of persons who respond correctly to item g. π_{gh} denotes the proportion of persons giving a correct response to both g and h. The covariance between the scores on items g and h is defined as $\sigma_{gh} = \pi_{gh} - \pi_g \pi_h$. Assuming that $g < h$ implies $\pi_g \leq \pi_h$ (π_g and π_h are fixed), the maximum possible covariance is obtained if $\pi_{gh} = \pi_g$; therefore, $\sigma_{gh}(\text{max}) = \pi_g(1 - \pi_h)$. Using these definitions, Loevinger's scalability coefficient H can be defined as (based on Mokken & Lewis, 1982; Sijtsma, 1988)

$$H = \sum_{g<h}\sum \sigma_{gh} \Big/ \sum_{g<h}\sum \sigma_{gh}(\text{max}) \quad . \tag{3}$$

This coefficient also can be written in terms of errors according to the Guttman (1950) scalogram (see Mokken, 1971, for the exact formula). In Equation 3, the maximum value of H (which is equal to 1) is obtained if $\sigma_{gh} = \sigma_{gh}(\text{max})$ for all item pairs; 0 is obtained if the average covariance equals 0; and a negative H value is obtained if the average covariance is negative.

Based on H, Mokken (1971) and Mokken and Lewis (1982) proposed an item scalability coefficient that only involves the covariances of a particular item with the other $k - 1$ items. For item g,

$$H_g = \sum_{h \neq g} \sigma_{gh} \Big/ \sum_{h \neq g} \sigma_{gh}(\text{max}) \quad . \tag{4}$$

[Refer to Mokken (1971) and Mokken and Lewis (1982) for all relevant definitions and proofs, as well as other results, with respect to H and H_g.]

For the application of H to the transposed data matrix (items $\times$ persons), let β_i denote the proportion of items to which person i gives the correct response. If a fixed test of k items is assumed, then β_i is the expectation of the observed proportion of correct responses across independent repeated measurements (see Lord & Novick, 1968, pp. 30–31, for a similar definition of the true score rather than the true proportion correct). In the sample, β_i is estimated by x_i/k, where x_i equals the number-correct score. Let β_{ij} denote the expected proportion of items to which both persons i and j respond correctly. In the sample, β_{ij} is estimated by x_{ij}/k, where x_{ij} equals the number of items to which both persons i and j respond correctly. $\sigma_{ij} = \beta_{ij} - \beta_i \beta_j$ is the covariance between the scores of persons i and j on k items. Assuming that $i < j$ implies $\beta_i \leq \beta_j$ (β_i and β_j are fixed), the maximum possible covariance is obtained if $\beta_{ij} = \beta_i$; therefore, $\sigma_{ij}(\text{max}) = \beta_i(1 - \beta_j)$. For a set of n persons, the H coefficient, now denoted as H^T, is defined as

$$H^T = \sum_{i<j}\sum \sigma_{ij} \Big/ \sum_{i<j}\sum \sigma_{ij}(\text{max}) \quad . \tag{5}$$

The corresponding coefficient for a particular person i is defined as

$$H_i^T = \sum_{j \neq i} \sigma_{ij} \Big/ \sum_{j \neq i} \sigma_{ij}(\text{max}) \quad . \tag{6}$$

Like H and H_g, H^T and H_i^T can be written in terms of error patterns. For an ability test, an error in the sense of the Guttman scalogram occurs if the more able person j answers an item, say g, incorrectly, and person i ($\beta_i \leq \beta_j$) answers item g correctly.

Obviously, all results for H and H_g hold by symmetry for H^T and H_i^T. The relation of H^T and H_i^T to the properties of the IRFs was of particular interest here. It is shown below that non-negative values of H^T and H_i^T are necessary conditions for a set of n person response functions (PRFs) to be nonincreasing, or equivalently, for a set of k IRFs to be nonintersecting. This is a new result in the

context of the Mokken approach. Cliff (1977) proposed a coefficient, denoted by c_{p3}, that is formally identical to H^T. The definition of c_{p3} is different from H^T, however. Also, its relation to the characteristics of PRFs or IRFs was not recognized.

The H^T Coefficient and Nonintersection of IRFs

Assume that given nonintersection of IRFs, for some value of θ success probabilities $\pi_g(\theta)$ ($g = 1, \ldots, k$) are strictly ordered so that

$$\pi_1(\theta) < \pi_2(\theta) < , \ldots , < \pi_k(\theta) \quad . \tag{7}$$

However, nonintersecting IRFs can touch locally; therefore, $\pi_g(\theta) = \pi_h(\theta)$ for some θs. IRFs may even coincide completely so that this equality holds for all θs. It must, therefore, be recognized explicitly that strict inequality (Equation 7) does not necessarily hold for each person: for some θs, ties may occur that are expressed by

$$\pi_1(\theta) \leq \pi_2(\theta) \leq , \ldots , \leq \pi_k(\theta) \quad . \tag{8}$$

Equation 8, in the absence of Equation 7, could be misleading because knowing that, for example, $\pi_g(\theta_i) = \pi_h(\theta_i)$, does not imply the strict order of these items for other θs.

Given the DM model and the ordering according to Equation 8, Mokken (1971, pp. 118, 126) defined the ordering of δ to be exactly the reverse (apart from ties):

$$\delta_1 \geq \delta_2 \geq , \ldots , \geq \delta_k \quad . \tag{9}$$

Although the DM model allows item discriminations and pseudochance levels to vary, the ordering of success probabilities in Equation 8 is completely determined by the ordering of the latent item difficulties in Equation 9. Using β notation, for an arbitrarily selected examinee i, this can be expressed as

$$\beta_i(\delta_1) \leq \beta_i(\delta_2) \leq , \ldots , \leq \beta_i(\delta_k) \quad . \tag{10}$$

The success probability $\beta(\delta)$ is a nonincreasing function of the item difficulty δ. Note that $\beta(\delta)$ can be interpreted as the PRF (Lumsden, 1978).

All PRFs $\beta(\delta)$ that have the same ordering are called similarly ordered functions (see Mokken, 1971, pp. 119–120, for a more general treatment of similarly ordered functions). For similarly ordered functions, say $\beta_i(\delta)$ and $\beta_j(\delta)$, the following definition applies for each pair of function values δ_g and δ_h:

$$[\beta_i(\delta_g) - \beta_i(\delta_h)][\beta_j(\delta_g) - \beta_j(\delta_h)] \geq 0 \quad . \tag{11}$$

Next, assume that a test consists of k items, each characterized by some value δ_g ($g = 1, \ldots, k$). δ may be either continuously or discretely distributed. Let $P(\delta)$ denote the cumulative distribution function of the item difficulty δ. Note that success probabilities are independent for different persons. Then it can be shown that for two similarly ordered functions in general (Mokken, 1971), and for $\beta_i(\delta)$ and $\beta_j(\delta)$ in particular,

$$\int_\delta \beta_i(\delta)\beta_j(\delta)dP(\delta) \geq \int_\delta \beta_i(\delta)dP(\delta)\int_\delta \beta_j(\delta)dP(\delta) \quad . \tag{12}$$

Integration across δ yields $\beta_{ij} \geq \beta_i\beta_j$, or in terms of the covariance between persons i and j,

$$\sigma_{ij} = \beta_{ij} - \beta_i\beta_j \geq 0 \quad . \tag{13}$$

Given nonzero variances for θ and δ, the maximum covariances in the denominator of H^T in Equation 5 are all positive constants; therefore, from Equation 13 it can be deduced that for nonincreasing PRFs, or equivalently, nonintersecting IRFs, $H^T \geq 0$. From this, the following theorem is obtained:

For a set of k items having nonintersecting IRFs (or equivalently, for n persons having nonincreasing PRFs),

$$0 \leq H^T \leq 1 \quad , \tag{14}$$

where $H^T = 0$ if and only if at least $n - 1$ PRFs $\beta(\delta)$ are constant over the region of δ where its density is not equal to 0.

The necessity and sufficiency of the condition for $H^T = 0$ follow from the general results presented by Mokken (1971) with respect to similarly ordered functions. Obviously, the theorem also holds for the person coefficient H_i^T.

A constant value of $\beta(\delta)$ for at least $n - 1$ θs means that for an arbitrary θ the success probability is the same on each of the k items. For different θs, these constant success probabilities need not be equal. In general, $\beta_i(\delta) \neq \beta_j(\delta)$ for $\theta_i \neq \theta_j$. The condition for $H^T = 0$ thus says that the IRFs of k items all have the same shape and the same location; in other words, all k IRFs coincide. An exception may occur for at most one θ value.

Practical Use of H^T

Interpretation of Numerical Values of H^T

Roskam et al. (1986) and Mokken et al. (1986) discussed H as a function of the interplay of the distance between the location parameters (item difficulties) of IRFs, the slope (discrimination power) of IRFs, and the variance of the person distribution. Given that IRFs are monotonically nondecreasing but are allowed to intersect, it is argued that H is an increasing function of each characteristic, if the other two are held constant. By symmetry, H^T is a function of the interplay of the distance between PRFs, the slope of the PRFs, and the variance of the item difficulties, δ. How these characteristics affect H^T is studied under the condition that PRFs are monotonically nonincreasing, but are allowed to intersect. Equivalently, it was thus assumed that IRFs do not intersect, but they do not necessarily have to be monotonically nondecreasing. Note that if the full DM model (i.e., the complete set of assumptions) were assumed, this last property also should have been assumed.

Given the condition that PRFs are monotonically nonincreasing but are allowed to intersect, if two characteristics are held constant H^T is an increasing function of the third. For example, keeping the distance and the slopes of the PRFs fixed, H^T increases as the variance of δ increases. If the variance of δ is 0, H^T is either a positive constant or equal to 0. As an extreme example, imagine a set of k nonintersecting IRFs that all touch at $\theta = \delta_1 = \delta_2 = , \ldots, = \delta_k$ but nowhere else. In this theoretical example, covariances between persons with $\theta \neq \delta$ are positive, implying $H^T > 0$. In general, for k nonintersecting IRFs that coincide at $\theta = \delta_1 = \delta_2 = , \ldots, = \delta_k$, a positive H^T is obtained if there are at least two θs for which at least two different success probabilities exist (i.e., for these θs not all IRFs coincide). Consequently, at least one positive covariance exists, so that $H^T > 0$. Only if all IRFs coincide, with the exception of at most one θ value (see theorem), will all the person covariances be equal to 0, implying that $H^T = 0$.

For example, for parallel IRFs (e.g., Rasch IRFs), if the variance of δ is 0, then $H^T = 0$. In this situation, a particular person would have the same success probability on the k items of the test. It would look as if each PRF is a constant function of δ, or equivalently, as if all IRFs coincide. It is clear that in this case success probabilities cannot be ordered. If the variance of δ increases, however, PRFs can be studied along a broader range. In this situation, H^T also increases. Thus, there is more

information available on the nonincreasingness of the PRFs, or equivalently, the nonintersection of the IRFs. As a result, more confidence can be held in the invariant ordering of the k success probabilities for different values of θ.

Thus, a negative H^T indicates a violation of the assumption of nonintersection of IRFs. Larger, positive H^T values provide more confidence in the invariant ordering of the success probabilities on k items.

The difference between H and H^T is particularly clear in the following situations. First, consider k coinciding IRFs that are all constant functions of θ. In this situation, $H = H^T = 0$. Second, suppose that none of the IRFs coincide, but all are constant functions of θ. This means that $H = 0$, but $H^T > 0$, and the further the IRFs are apart (the steeper the PRFs are), the larger H^T will be. Third, suppose that the IRFs coincide, but all are increasing functions of θ. This means that $H^T = 0$, but $H > 0$, and the steeper the IRFs are, the larger H will be.

A Practical Lower Bound for H^T

A negative H^T value is probably rare in practice. Furthermore, a non-negative H^T does not exclude intersection of IRFs. Analogously, a negative H value for items is rare, and a positive H value does not exclude that some IRFs are (locally) decreasing. Mokken (1971, p. 184) and Mokken and Lewis (1982) therefore proposed that a set of items satisfy the following requirements for a suitably selected constant c:

$$\sigma_{gh} > 0 \quad , \tag{15}$$

for all item pairs; and

$$H_g \geq c > 0 \quad , \tag{16}$$

for all items.

Because $H \geq \min(H_g)$ (Mokken & Lewis, 1982), the second requirement implies $H \geq c > 0$. Mokken recommended selecting $c = .3$; this lower bound excludes most violations of the property of nondecreasingness of IRFs. Items violating these requirements are removed from the analysis.

For studying the nonintersection of IRFs, substituting person covariances (σ_{ij}) and person coefficients (H_i^T) in Equations 15 and 16, respectively, would be the incorrect procedure to follow for two reasons. First, from a practical point of view, the inspection of all person covariances is both laborious [$\binom{n}{2}$ covariances] and an insecure strategy (each covariance is based on only k observations). Second, removal of "outlying" persons ($H_i^T < c$) from the analysis would yield a biased sample with an artificially high H^T estimate.

Therefore, for H^T and H_i^T alternative requirements are proposed that are deduced from a monte carlo study discussed below:

$$H^T \geq a \quad , \tag{17}$$

percentage of negative $H_i^T < b \quad , \tag{18}$

where a and b are positive constants. Given such values, it may be expected that if IRFs intersect, but H^T nevertheless has a value of at least a, several (preferably at least $b\%$) H_i^T coefficients will have negative values, thereby correctly indicating intersection.

Sampling Characteristics of H^T

Mokken (1971, pp. 157-169; also see Goodman, 1959) provided the asymptotic sampling theory

for the null hypotheses that $H = 0$ and that two or more H coefficients for the same test in different populations are equal, respectively. Application of this sampling theory to H^T would be problematic because the sample of items is always very small. A serious discrepancy between the empirical and the theoretical distribution of H^T probably exists. Further, it would have to be assumed that samples of k items are randomly drawn from a population of items, which does not seem to be realistic in most applications. As an alternative, characteristics of the sampling distribution of H^T were studied by means of a monte carlo strategy.

A Monte Carlo Study

Method

Different levels of violation of the property of nonintersecting IRFs were defined by different configurations of a fixed number of IRFs. Note, in particular, that levels of violation do not exist independently of the characteristics of IRFs, but rather are defined by them. Thus, taking three sets of equidistant parallel IRFs (no violations) with different spacing across sets as a starting point, one or more IRFs were replaced by flatter IRFs (to create more intersections). Combined with two levels of different overall discrimination, a completely crossed design was constructed, each cell representing a level of violation.

Specifically, data matrices (100 persons $\times$ 9 items) containing artificial binary scores were obtained (for the simulation procedure see Sijtsma & Molenaar, 1987) for two-parameter logistic IRFs and a standard normal distribution of θ. A $2 \times 3 \times 5$ design was considered: two levels of overall discrimination ($\alpha = 1, 2$), three levels of distance between item difficulties [$d(\delta) = 0.0, .2, .5$; the δs were equidistant with median 0], and five levels of violation of the property of nonintersecting IRFs (one level was the case of no violations, the other four pertained to increasing degrees of violation). For each cell, the following data were computed: (1) the value of H^T, averaged over 100 random samples; (2) the standard deviation (SD) of H^T, taken across 100 random samples; and (3) the average percentage of negative H_i^T values.

Results

As the theoretical considerations predict, an increase in either distance [$d(\delta)$] or discrimination power (α) resulted in an increasing value of H^T (see Table 1). For $d(\delta) > 0$, H^T decreased as the seriousness of the violation of the property of nonintersection of IRFs increased. For $d(\delta) = 0$ in Case 1 (no intersection; coinciding IRFs) $H^T \approx 0$ ($H^T = 0$ is stipulated by the theorem). For Cases 2 through 5, there were always two subsets of coinciding IRFs that had their cutting point at $\delta = 0$, which was in the middle of the symmetric person distribution. As a result, theoretically there were only two (reverse) partial orderings of success probabilities that are equally represented. This condition yields $H^T = 0$.

Because H^T is a positively weighted average of $H_i^T (i = 1, \ldots, k)$ (see Mokken & Lewis, 1982, for the relationship between H and H_g, $g = 1, \ldots, k$), the percentage of negative H_i^T values tended to increase as H^T decreased (Table 1). The SD of H^T was approximately .01 for all cases in which (subsets of) IRFs coincided (with or without intersections). For IRFs that did not coincide for items [$d(\delta) > 0$] and that either did or did not intersect, the SD varied from approximately .02 to .05.

For cells with $d(\delta) > 0$, using $n = 200$ rather than $n = 100$ yielded only minor changes in the mean H^T: the largest absolute difference ($n = 100$) was .017. The SD showed an average decrease of .01 with the larger n. With test length $k = 18$ ($n = 100$), the overall results for H^T and the SD remained similar to those obtained with $k = 9$. Specifically, the SD ranged from .02 to .03, which is, on the

Table 1
Mean and Standard Deviation (SD) of H^T, and Mean Percentage of Negative H_i^T Values ($\%H_i^T$) for Five Cases and Items of Varying Discrimination ($n = 100$, $k = 9$, and 100 Replications Per Cell)

Case and Item Discriminations	$d(\delta) = 0.0$			$d(\delta) = .2$			$d(\delta) = .5$		
	Mean	SD	$\%H_i^T$	Mean	SD	$\%H_i^T$	Mean	SD	$\%H_i^T$
Weak Discrimination									
1 (1,1,1,1,1,1,1,1,1)	.001	.010	49.6	.102	.030	24.6	.402	.042	4.8
2 (1,1,1,1,1,1,1,1,.5)	0.000	.009	49.2	.080	.025	27.7	.360	.038	6.0
3 (1,1,1,1,1,1,1,.5,.5)	−.001	.007	50.4	.068	.024	28.7	.312	.036	7.5
4 (1,1,1,1,1,.5,.5,.5,.5)	0.000	.007	48.4	.059	.023	30.3	.274	.041	9.4
5 (1,1,1,1,.5,.5,.5,.5,.5)	0.000	.008	49.9	.057	.023	30.6	.265	.040	9.5
Strong Discrimination									
1 (2,2,2,2,2,2,2,2,2)	−.001	.013	50.0	.316	.041	9.8	.743	.035	.3
2 (2,2,2,2,2,2,2,2,.5)	−.001	.012	49.8	.220	.045	14.6	.662	.035	.8
3 (2,2,2,2,2,2,2,.5,.5)	.001	.011	48.9	.156	.032	19.6	.543	.038	1.8
4 (2,2,2,2,2,.5,.5,.5,.5)	0.000	.009	49.4	.113	.030	24.0	.420	.042	4.1
5 (2,2,2,2,.5,.5,.5,.5,.5)	−.001	.009	50.1	.099	.027	25.3	.391	.036	4.5

average, somewhat smaller than for $k = 9$. Furthermore, the conclusions concerning the percentage of negative H_i^T values were not affected by a larger sample size or a longer test.

Discussion and Conclusions

Given these results, tentative values for a and b are proposed: $H^T \geq .3$ and the percentage of negative H_i^T values < 10. Thus, if $H^T \geq .3$ and the percentage of negative H^T values < 10, it is assumed for all practical purposes that the IRFs do not intersect. If one or both of these requirements are not satisfied, it is assumed that the item ordering is different for substantial numbers of persons. Experience might lead to the conclusion that other values or more complex rules are more appropriate.

The cells in Table 1 for $d(\delta) = .5$, $\alpha = 2$, and Cases 2 through 5, which reflect modest to strong violations of the property of nonintersection of IRFs illustrate that the proposed bounds do not always lead to correct decisions about nonintersection of IRFs. Note, however, that using the Mokken approach for item analysis removes the weakly discriminating items ($H_g < .3$) from the analysis prior to investigation of the nonintersection of IRFs. Consequently, in practice, H^T is applied to relatively strongly discriminating item sets.

The following methodology for using H^T is proposed. First, the more general MH model is investigated by inspection of the covariances between the items, the H_g values, and an additional method (Meijer et al., 1990; Molenaar, 1982, 1983) that is based on the model property that the proportion of correct responses on item g is a nondecreasing function of the "rest" score (i.e., the total score minus the score on item g). Second, given the MH model, nonintersection of IRFs is investigated by examining the H^T coefficient and the percentage of negative H_i^T values. Given a negative outcome, visual inspection methods proposed by Mokken (1971; also see Meijer et al., 1990) can be used to find deviant items. After removing the deviant items, whether $H^T \geq .3$ and the percentage of negative H_i^T values < 10 should again be evaluated.

References

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1985). *Revisie Amsterdamse Kinder-Intelligentie Test (RAKIT)* [Revision of the Amsterdam Child Intelligence Test]. Lisse: Swets & Zeitlinger.

Cliff, N. (1977). A theory of consistency of ordering generalizable to tailored testing. *Psychometrika, 42,* 375–401.

Goodman, L. A. (1959). Simple statistical methods for scalogram analysis. *Psychometrika, 24,* 29–43.

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

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton: Princeton University Press.

Holland, P. W. (1981). When are item response models consistent with observed data? *Psychometrika, 46,* 79–92.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14,* 1523–1543.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory.* Homewood IL: Dow Jones-Irwin.

Kingma, J., & TenVergert, E. M. (1985). A nonparametric scale analysis of the development of conservation. *Applied Psychological Measurement, 9,* 375–387.

Lewis, C. (1983). Bayesian inference for latent abilities. In S. B. Anderson & J. S. Helmick (Eds.), *On Educational Testing* (pp. 224–251). San Francisco: Jossey-Bass.

Loevinger, J. (1948). The technique of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychological Bulletin, 45,* 507–530.

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

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading MA: Addison-Wesley.

Lumsden, J. (1978). Tests are perfectly reliable. *British Journal of Mathematical and Statistical Psychology, 31,* 19–26.

Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement, 14,* 283–298.

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* New York/Berlin: De Gruyter.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6,* 417–430.

Mokken, R. J., Lewis, C., & Sijtsma, K. (1986). Rejoinder to "The Mokken scale: A critical discussion." *Applied Psychological Measurement, 10,* 279–285.

Molenaar, I. W. (1982). Een tweede weging van de Mokken-schaal [A second weighing of the Mokken scaling procedure]. *Tijdschrift voor Onderwijsresearch, 7,* 172–181.

Molenaar, I. W. (1983). Rasch, Mokken en schoolbeleving [Rasch, Mokken, and school experience]. In S. Lindenberg & F. N. Stokman (Eds.), *Modellen in de Sociologie* (pp. 195–213). Deventer: Van Loghum Slaterus.

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49,* 425–435.

Rosenbaum, P. R. (1987a). Comparing item characteristic curves. *Psychometrika, 52,* 217–233.

Rosenbaum, P. R. (1987b). Probability inequalities for latent scales. *British Journal of Mathematical and Statistical Psychology, 40,* 157–168.

Roskam, E. E., van den Wollenberg, A. L., & Jansen, P. G. W. (1986). The Mokken scale: A critical discussion. *Applied Psychological Measurement, 10,* 265–277.

Sijtsma, K. (1988). *Contributions to Mokken's nonparametric item response theory.* Amsterdam: Free University Press.

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika, 52,* 79–97.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52,* 589–617.

Author's Address

Send requests for reprints or further information to Klaas Sijtsma, Vakgroep Arbeids en Organisatiepsychologie, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.