

A New Approach to Desirable Responding: Multidimensional Item Response Model of Overclaiming Data

Applied Psychological Measurement
2023, Vol. 47(3) 221–236

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01466216231151704

journals.sagepub.com/home/apm



Kuan-Yu Jin¹ , Delroy L. Paulhus², and Ching-Lin Shih^{3,2} 

Abstract

A variety of approaches have been presented for assessing desirable responding in self-report measures. Among them, the overclaiming technique asks respondents to rate their familiarity with a large set of real and nonexistent items (foils). The application of signal detection formulas to the endorsement rates of real items and foils yields indices of (a) *knowledge accuracy* and (b) *knowledge bias*. This overclaiming technique reflects both cognitive ability and personality. Here, we develop an alternative measurement model based on multidimensional item response theory (MIRT). We report three studies demonstrating this new model's capacity to analyze overclaiming data. First, a simulation study illustrates that MIRT and signal detection theory yield comparable indices of accuracy and bias—although MIRT provides important additional information. Two empirical examples—one based on mathematical terms and one based on Chinese idioms—are then elaborated. Together, they demonstrate the utility of this new approach for group comparisons and item selection. The implications of this research are illustrated and discussed.

Keywords

knowledge exaggeration, overclaiming, signal detection theory, multidimensional item response theory, programme for international student assessment, Chinese idiom

Introduction

Despite their limitations, self-report measures have continued to be a mainstream method for human assessment in educational, psychological, and applied settings (Paulhus & Vazire, 2007). Their popularity stems from several advantages: they are easy to administer and inexpensive to

¹Hong Kong Examinations and Assessment Authority, Hong Kong, Hong Kong

²University of British Columbia, Vancouver, BC, Canada

³National Sun Yat-Sen University, Kaohsiung, Taiwan

Corresponding Author:

Ching-Lin Shih, Institute of Education and Center for Teacher Education, Assessment Research Center, National Sun Yat-sen University, No. 70, Lienhai Rd., Kaohsiung 80424, Taiwan.

Email: educls@mail.nsysu.edu.tw

score compared to informant or behavioral alternatives. In addition, they can be completed within a reasonable short time, facilitating the recruitment of survey participants.

In spite of these advantages, self-report measures have long drawn concerns about response styles (Cronbach, 1949). Arguably, the most threatening of these response styles is socially desirable responding (SDR; Paulhus, 1991). Differential tendencies to give desirable responses can contaminate valid responses. The impact of SDR on self-report measures of personality remains controversial despite 80 years of empirical investigation (for a review, see Ziegler et al., 2011). Because of this potential threat, it is important to distinguish the valid content of self-reports from the distortion engendered by SDR. This goal requires measurement of SDR. However, the most tempting solution—partialing SDR out of self-reports—has proven counterproductive (Paulhus, 2017). Advocates for the superiority of cognitive ability measures point out that, by contrast, their scoring is objective; that is, responses are valid if they match a scoring key.

To overcome the limitations of self-reports, objective performance methods are generally preferred (Furr & Funder, 2007). In the literature, proactive and reactive methods have been applied to measure or control SDR. Proactive methods are used to enhance respondents' motivation to provide true answers before data are collected, whereas reactive methods are used to detect or correct biased responses after data are collected. One may refer to Muszyński (2020) for a detailed review and comparison of different methods.

Among them, two reactive methods, which are relatively easy to implement in practice, are briefly introduced here (Paulhus & Holden, 2010). One exploits the fact that faking a response takes longer than responding honestly (Holden et al., 1992). Thus, Holden and Kroner (1992) were able to diagnose SDR by comparing response times under faked versus honest instructions.

A second objective method, called the *overclaiming technique* (OCT; Paulhus et al., 2003), requests respondents to indicate their level of familiarity with a set of items belonging to a specific content domain (e.g., historical events, famous musicians). To address exaggeration, the item set included some items that do not exist (called *foils*). Any endorsement of such foils is of particular interest where there might be a cognitive bias (Müller & Moshagen, 2018) or a tendency of self-enhancement (Paulhus et al., 2003). In this study, such endorsements were taken as representing a blatant misrepresentation of reality. Indeed, this concept of overclaiming bias has been designated as one of the “five pillars of self-enhancement” (Sedikides & Alicke, 2019).

Here, we use the term “overclaiming data” for any dataset comprising responses to a combination of real items and foils. Because such datasets incorporate a key for “truth” (i.e., the factual difference between reals and foils), the accuracy of an individual's responses can be scored without any external criterion and can be taken as an objective performance measure (Paulhus & Harms, 2004). Although presented as a nonthreatening survey of everyday familiarities, the OCT method can be used to tap both *knowledge accuracy* (i.e., true familiarity with knowledge) and *knowledge bias* (i.e., knowledge exaggeration) for each respondent (Paulhus, 2011).

For example, the academic version, the Over-Claiming Questionnaire-150 (OCQ-150; Paulhus et al., 2003), covered 10 content domains (e.g., historical names and events, fine arts, language, and popular culture), each of which contained 15 items. In that initial version, the real and foil items were 80% and 20%, respectively. Respondents were asked to rate their familiarity with each item on a 7-point scale ranging from 0 (never heard of it) to 6 (know it very well). Because the 150-item version is rather lengthy to apply widely in practice, Paulhus et al. (2003) shortened the instrument to 60 items. Other studies have used even briefer measures (e.g., Bing et al., 2011) and alternative content (Jones et al., 2020; Paulhus & Dubois, 2014; Ziegler, Kemper, & Rammstedt, 2013). The standard analytical approach of OCT was derived from signal detection theory (SDT; Swets, 1964). Under that terminology, the endorsements of reals and foils are labeled *hits* and *false alarms*, respectively. Using a variety of standard

formulas (MacMillan & Creelman, 2005), those endorsement rates can then be transformed to capture *knowledge accuracy* and *knowledge bias* (Paulhus et al., 2003; see details below). A more elaborate method requires confirmation of hits (Ackerman & Ellingsen, 2014). Finally, a structural equation alternative has been recommended by Hülür et al. (2011).

Item Response Theory

Item response theory (IRT) has become the predominant measurement model in many fields. For example, its advantages have now been exploited in educational (e.g., Wang et al., 2006) and personality research (e.g., Reise & Henson, 2003). In applying IRT to overclaiming data, Williams and Zumbo (2003) had to confront a challenge for any alternative to the SDT; namely, that accuracy scores require both reals and foils. They overcame this challenge by analyzing a 75-item overclaiming dataset using a two-stage method. The first step was to obtain a latent accuracy factor score for each respondent. This goal was effected by creating 12 groups of “pseudo items” based on the subscale structure. The accuracy score of each pseudo item was calculated from the averaged hit rates (for real items) and false alarm rates (for foils). A factor analysis was then conducted on the 12 accuracy scores, with the factor scores being used to represent the latent variable. The second step was to derive an item characteristic curve (ICC) for each pseudo item (i.e., each subscale). To do so, the locally weighted scatterplot smoothing method (Fox, 1997) was fitted to each scatterplot: The *x*- and *y*-axes represented the latent variable and the standardized accuracy score on each pseudo item, respectively. Thus, the obtained nonparametric ICCs represented the functioning of pseudo items on the latent variable.

Although novel, the Williams–Zumbo approach has limited value for analyzing overclaiming data. First, it deals only with accuracy, overlooking knowledge bias entirely. Second, the approach does not address raw items directly; rather, the pseudo items are at the subscale level (i.e., topics). Consequently, the relationship between the derived latent variable and the raw items remains unclear. Finally, because an accuracy score is averaged over the raw items within the pseudo item, the selection of raw items highly influences the quality of the pseudo item. This may be an issue for subscales with few raw items. In addition, a substantial number of pseudo items are required to satisfy the assumptions of factor analysis. Hence, this method is most appropriate for a longer questionnaire composed of many subscales. Recently, SDT indices have been to be problematic and less straightforward in representing two independent latent variables, as was commonly assumed (Goecke et al., 2020). To address these limitations, we propose a multidimensional item response theory (MIRT) framework to model overclaiming data directly at the item level.

The present article is organized as follows. First, formulas for the SDT approach are reviewed, and formulas for the MIRT approach are introduced. Then, parallels between the SDT indices and the MIRT model parameters are highlighted. Next, we detail a simulation study and two empirical applications to demonstrate the capability of this new method. Finally, the results and suggestions for further research are discussed.

Applying Signal Detection Theory (SDT)

As noted earlier, the SDT approach has been the standard method for OCT data analysis. First introduced by Swets (1964), SDT classifies the response to each item into one of four categories: (1) hit, (2) miss, (3) false alarm, or (4) correct rejection. A *hit* and a *miss* refer to the cases where an existent stimulus has or has not been correctly recognized by the respondent, respectively; a *false alarm* refers to a nonexistent stimulus that has been incorrectly claimed, and a *correct rejection* indicates an accurate denial of receiving a nonexistent stimulus.

In SDT, a respondent's judgment as to whether a signal is presented or not (i.e., a stimulus is real or not) can be depicted on a latent continuum of familiarity with two standardized normal distributions (DeCarlo, 2020; 2021). Figure 1 illustrates the formulation visually. In the panels of Figure 1, the right and left distributions represent respondents' familiarity for real items and foils, respectively. Figure 1(a) shows the case in which a real stimulus is presented (i.e., the distribution plotted with solid line). Compared to the conceptual distribution (graphed with a dashed line on the left) for respondents who do not recognize the real stimulus, the areas above and below the criterion refer to hit and miss rates, respectively. Figure 1(b) shows the case in which a foil is presented (i.e., the left distribution with solid line). Compared to the conceptual distribution (plotted with dashed line on the right) for respondents who deem the stimulus to be real, the areas above and below the criterion refer to false alarm and correct rejection rates, respectively.

In the present case, all items (reals and foils) are treated as homogeneous; only the differences between subjects were of interest here. As indicated in Figure 1, two person parameters are associated with the probabilities of the four response categories (hit, miss, false alarm, and correct rejection). First, knowledge accuracy for respondent n , which is defined as the number of hits relative to the number of false alarms, is calculated and standardized as

$$d'_n = z(H_n) - z(F_n), \quad (1)$$

where $z(H_n)$ and $z(F_n)$ are the z -transformed hit and false alarm rates of respondent n , respectively. A respondent with a higher d'_n shows greater accuracy in discriminating between existent and nonexistent stimuli.

The criterion c_n , which is known as the *knowledge bias index* (MacMillan & Creelman, 2005; Paulhus et al., 2003; Sorkin, 1999), is used to represent the respondent n 's tendency to overclaim. It is calculated as follows

$$c_n = -\frac{z(H_n) + z(F_n)}{2}, \quad (2)$$

where c_n equals zero only if the false alarm rate is equal to the miss rate, i.e., $z(1 - H_n) = -z(H_n) = z(F_n)$. When the false alarm rate is greater than the miss rate, that is, $z(F_n) > -z(H_n)$, a negative c_n is obtained. However, c_n becomes positive when the false alarm rate is lower than the miss rate. The lower the c_n value (i.e., the farther to the left that the criterion shifts along the latent continuum; for example, Figure 1(b)), the higher the tendency to overclaim. For a more intuitive interpretation¹, we define $\hat{c}_n = -c_n$ hereafter with the label \hat{c}_n corresponding to the mean of the hit rate and the false alarm rate.

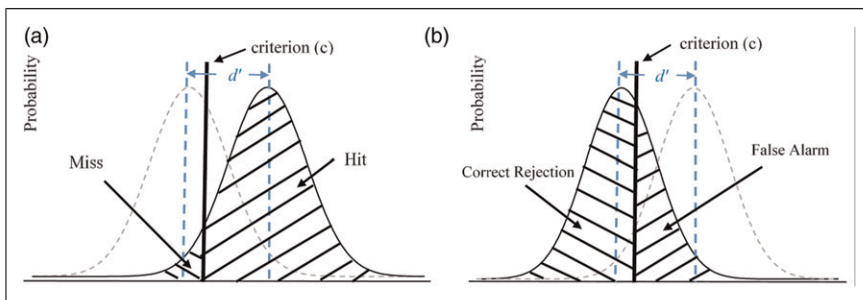


Figure 1. The Signal Detection Model for (a) real items with corresponding hit and miss rates and (b) foils with corresponding false alarm and correct rejection rates.

The SDT approach is not limited to binary outcome variables; it can also be extended to accommodate ordinal (rating) responses (MacMillan & Creelman, 2005). Consider a set of items with J categories (scored from 0 to $J - 1$) and scores of 1 to $J - 1$ used sequentially as cutoff values. For example, the first cutoff was set at 1, which means that every response scored at 1 or above was taken as a hit (for a real item) or false alarm (for a foil), and the d' and c' indices were in turn calculated for each respondent. These procedures were then repeated for the remaining $J - 2$ cutoffs, and two indices were computed accordingly. The overall accuracy and bias score for each respondent were obtained by averaging the $J - 1$ d' and c' indices.

Although SDT can be used to characterize individual differences in knowledge accuracy and bias (Paulhus et al., 2003), a complication ensues from the fact that each item response is actually a function of two elements: *personal features* and *item features*. As Williams and Zumbo (2003) correctly noted, the measurement of item features remained unclear in the SDT approach. To accurately predict a person's response to an item and correctly evaluate the measurement quality, both personal and item features should be jointly modeled. Given this more precise representation of each person's response to each item, the extent of a respondent's misrepresentation (bias) can also be recovered more cleanly.

Also, in SDT analyses, the personal indices are typically reported without measurement errors. This deficiency precludes a comparison between the two estimates of an individual's accuracy and distracts from possible inaccuracies in decision-making about an individual's accuracy (or bias). Third, although the two indices are useful for determining the relative location of a subject within a sample, the practical implications are not evident. The indices have no units and therefore require further explication. To address these drawbacks, we now follow-up the Williams and Zumbo (2003) method and promote an alternative: the MIRT approach.

Applying Multidimensional Item Response Theory (MIRT)

Self-report inventories, including overclaiming questionnaires, are usually composed of items rated on an ordinal scale. Over the past several decades, a number of studies have developed or used different IRT models to accommodate ordinal (rating) data under various conditions (Jeon & De Boeck, 2016; Jin & Wang, 2014; Tay et al., 2011; Wang et al., 2006). In the present article, we adopted an IRT measurement model capable of dealing simultaneously with overclaiming data (i.e., a set of reals and foils). Similar to SDT analysis, two underlying assumptions were made as follows: first, the operation of knowledge bias inflates the scores of both reals and foils; and second, the knowledge accuracy reflects not only the endorsement of reals but also the correct rejection of foils. The anticipated *multidimensional model for overclaiming* (MMOC) can then be formulated as follows

$$P(Y_{ni} = 1) = \frac{\exp(\lambda_{i1}\theta_n + \lambda_{i2}\xi_n - \delta_i)}{1 + \exp(\lambda_{i1}\theta_n + \lambda_{i2}\xi_n - \delta_i)} \quad (3)$$

$$P(Y_{ni} = 1) = \frac{\exp(\lambda_{i1}\theta_n - \lambda_{i2}\xi_n - \delta_i)}{1 + \exp(\lambda_{i1}\theta_n - \lambda_{i2}\xi_n - \delta_i)} \quad (4)$$

where $P(Y_{ni} = 1)$ represents the probability that respondent n endorses item i , and θ_n and ξ_n represent the tendency of person n to exaggerate (i.e., knowledge bias) and his/her actual familiarity with the items (i.e., knowledge accuracy), respectively. Parameters λ_{i1} and λ_{i2} are the slopes of the two latent variables, and δ_i is the difficulty of item i . For the purpose of model identification, θ_n and ξ_n are constrained to follow an orthogonal bivariate normal distribution with

a mean vector of $\mu' = [0, 0]$ and a variance–covariance matrix of $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ in this study.

Figure 2 presents the two-dimensional item characteristic curves for reals and foils. This illustrates that, for a given accuracy level, a stronger bias induces a higher probability of endorsing both reals and foils. On the other hand, for a given bias level, higher accuracy will increase the probability of endorsing a real item and decrease that of endorsing a foil. The ratio of λ_{i1} to λ_{i2} , defined as *bias saturation*, indicates the degree of overclaiming on item i (for similar precedents, see Wang et al., 2015). The larger the ratio, the more substantial the impact of knowledge bias. The numerator and denominator are restricted to positive numbers, so the ratios are always positive.

Equations (3) and (4) can be easily generalized to fit ordinal data. For example, the probability of endorsing category y on item i of person n can be expressed as

$$P(Y_{ni} = y) = \frac{\exp \sum_{j=0}^y (\lambda_{i1} \theta_n + \lambda_{i2} \xi_n - \delta_i - \tau_{ij})}{\sum_{k=0}^{J-1} \exp \sum_{j=0}^k (\lambda_{i1} \theta_n + \lambda_{i2} \xi_n - \delta_i - \tau_{ij})} \quad (5)$$

$$P(Y_{ni} = y) = \frac{\exp \sum_{j=0}^y (\lambda_{i1} \theta_n - \lambda_{i2} \xi_n - \delta_i - \tau_{ij})}{\sum_{k=0}^{J-1} \exp \sum_{j=0}^k (\lambda_{i1} \theta_n - \lambda_{i2} \xi_n - \delta_i - \tau_{ij})} \quad (6)$$

where τ_{ij} is the j -th category threshold of item i and $\tau_{i0} \equiv 0$ by definition, and the other variables are defined as in Equations (3) and (4).

The limitations of using the SDT model to analyze overclaiming data, as mentioned above, can be easily addressed by the MMOC. The item response function of the MMOC is helpful for describing the relationship between person features, item features, and observed responses. Moreover, measurement errors are available for both θ and ξ ; these are necessary for estimating test reliability. Precise item response prediction is feasible by simultaneously applying the person and item estimates. These values can also be applied to evaluate the misfit of response vectors against the measurement model's expectation or to predict individual responses to unanswered items.

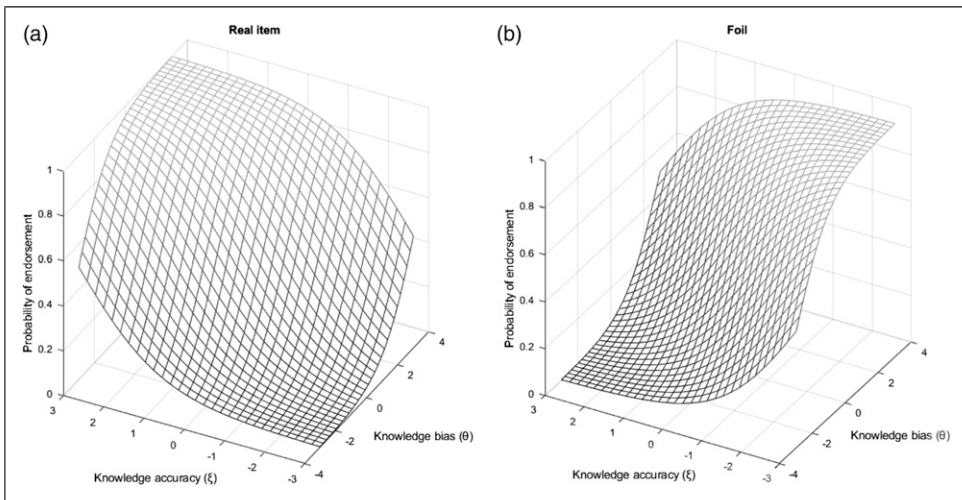


Figure 2. Item characteristic curves for real items and foils.

Of special importance, item parameters can be obtained when the MMOC is applied to overclaiming data. This information may be useful in practice (e.g., pruning items for a shorter version of the OCQ). Hence, the quality of measurement can be further ensured. As for parameter estimation, because the MMOC is a restricted model of the multidimensional random coefficients multinomial logistic model (Adams et al., 1997), many existing computer programs can feasibly be used for parameter calibration. We used the R package, TAM (Kiefer et al., 2020), implementing the *tam.mml* function with marginal maximum likelihood estimation.

Relations Between the SDT and MIRT Analyses of Overclaiming Data

In the SDT model, the hit rate is directly proportional to d' and c' , whereas the false alarm rate is directly proportional to c' and inversely proportional to d' . These relationships also hold in the MMOC. Specifically, the expected hit rate for combinations of θ and ξ is calculated by averaging the probability of positive responses across all reals. Thus, a higher θ (or ξ) would lead to a higher hit rate. Similarly, the expected false alarm rate is calculated by averaging the probability of positive responses across all foils; a higher θ (or a lower ξ) would therefore lead to a higher false alarm rate. In sum, the SDT and the MMOC approaches are conceptually equivalent: the SDT accuracy index, d' , is analogous to the ξ parameter in the MMOC, and the SDT bias index, c' , is analogous to the θ parameter. Next, we demonstrate the empirical relationship between these two approaches in a simulation study, after which we present two empirical examples. Due to space limitations, the second example is included in the [Appendix A](#).

Method

Simulation Study

A systematic series of simulations was conducted to compare the accuracy and bias indices obtained from the MMOC with those obtained from SDT. In particular, (a) correlations were evaluated between the two methods of calculating each index, and (b) the values of these correlations were investigated as a function of several design factors. Notably, because SDT is the most common procedure for analyzing overclaiming data (e.g., Kam et al., 2015; Müller & Moshagen, 2019; Paulhus et al., 2003; Vonkova et al., 2018), it was treated as a comparative reference for the proposed MMOC in this study.

Design

For each of the 48 experimental conditions, we generated datasets of 1000 hypothetical respondents. This process was then replicated 200 times. The results under each condition were then represented as the means of the 200 replications. The 48-condition design emerged from manipulating three factors: (1) the test length (10, 20, or 30 items); (2) the proportion of foils (20%, 30%, 40%, or 50%); and (3) the number of response categories (2, 3, 4, or 5). The person parameters (θ and ξ) were drawn from a multivariate normal distribution with mean $\mu' = [0, 0]$ and covariance $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The item parameters were drawn from the following distributions: $\lambda_{i1} \sim \text{Lognormal}(0, 0.1)$, $\lambda_{i2} \sim \text{Lognormal}(0, 0.1)$, and $\delta_{ij} \sim \text{Uniform}(-3, 3)$, with $\delta_{ij} < \delta_{ij'}$ for $j < j'$. Given the person and item parameters, the probabilities of endorsing each response category were given by [Equations \(3\)–\(6\)](#).

Therefore, the cumulative probability of a person's response falling in or below category j can be calculated, and it was then compared to a random number x sampled from *Uniform*(0, 1). The response of person n on item i was set to j , while x fell between the cumulative probabilities for categories $j - 1$ and j .

Analyses

The simulated data were analyzed with the R package TAM. The R script for calculating the MMOC indices from the data is provided in the [Appendix B](#). The corresponding SDT indices were similarly calculated with R. To determine the similarity of person indices from the SDT and MIRT, two dependent variables were calculated: the correlation between θ and c' (denoted by $r_{\theta c'}$), and the correlation between ξ and d' (denoted by $r_{\xi d'}$). The higher the correlation coefficients, the more interchangeable the MIRT and SDT person indices are. In addition, to investigate the impact of the manipulated factors on the two dependent variables, two sets of three-way ANOVAs were conducted. Because correlation coefficients (ranging from -1 to 1) violate the normality assumption of ANOVA, a Fisher z -transformation was applied to create approximately normally distributed dependent variables ($z_{\theta c'}$ and $z_{\xi d'}$) for analysis.

Results

Figure 3 shows the distributions of $r_{\theta c'}$ and $r_{\xi d'}$ across conditions. These values were higher than .70 for most conditions, supporting the proposition that the indices of knowledge accuracy and knowledge bias are directly related to the two latent trait estimates in the MMOC. However, when the dataset dipped to 10 dichotomous items, the correlation coefficients were less impressive.

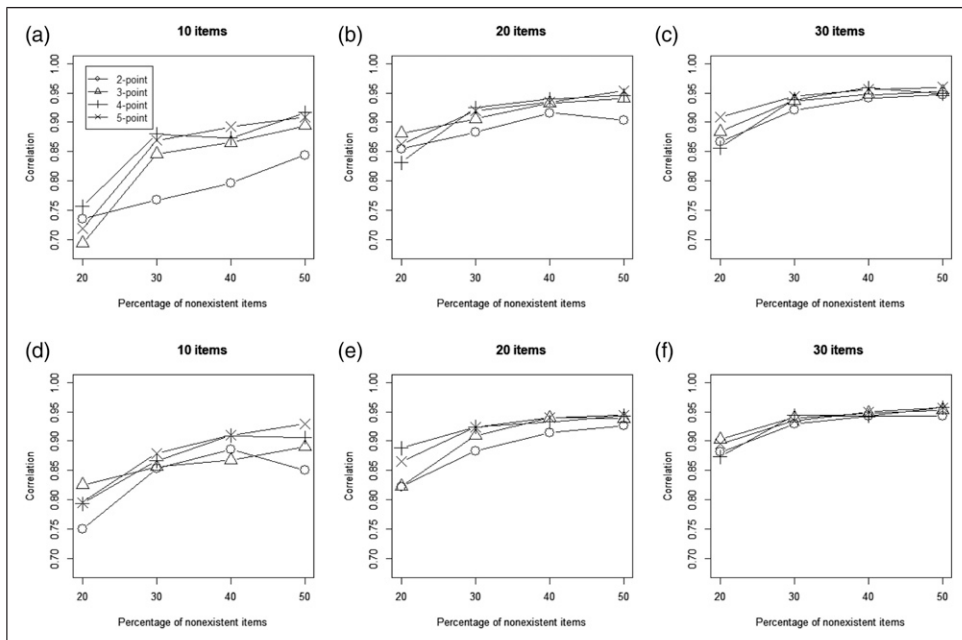


Figure 3. Relationships between the SDT and MMOC indices across conditions.

Note. The first row (a, b, c) refers to knowledge bias; the second row (d, e, f) refers to knowledge accuracy.

Table 1. Results of the Three-way ANOVA on $z_{\theta c'}$ and $z_{\xi d'}$.

Effects	SS	df	MS	p-value	Partial η^2
Dependent variable: $z_{\theta c'}$					
Test length (A)	374.245	2	187.122	<.001	.978
% of foils (B)	51.627	3	17.209	<.001	.862
Response category (C)	320.921	3	106.974	<.001	.975
A \times B	7.454	6	1.242	<.001	.474
B \times C	6.096	6	1.016	<.001	.425
A \times C	15.466	9	1.718	<.001	.652
A \times B \times C	16.064	18	0.892	<.001	.660
Error	4.044	9552	0.001		
Dependent variable: $z_{\xi d'}$					
Test length (A)	228.153	2	114.077	<.001	.965
% of foils (B)	30.525	3	10.175	<.001	.786
Response category (C)	267.642	3	89.214	<.001	.970
A \times B	5.966	6	0.994	<.001	.418
B \times C	2.484	6	0.414	<.001	.231
A \times C	3.967	9	0.441	<.001	.324
A \times B \times C	17.674	18	0.982	<.001	.681
Error	8.291	9552	0.001		

Note. $R^2 = .990$ in first analysis and $R^2 = .985$ in the second analysis.

Differences in correlations across foil proportions were also most discrepant under the 10-item condition. All correlations were found to consistently range from .85 to .95 under the 30-item condition.

Table 1 shows the results of two sets of three-way ANOVAs. The partial η^2 for the main effects and interactions ranged from .425 to .978 and from .231 to .970 for the z -transformed variables $z_{\theta c'}$ and $z_{\xi d'}$, respectively. These results indicate that the three independent variables were all crucial factors in determining the extent of correspondence between the SDT and MIRT approaches: a longer test, a larger proportion of foils, and more response categories were all helpful for enhancing the interchangeability of the SDT and MMOC indices. In short, the θ -parameter in the MMOC was successfully mapped onto the SDT knowledge bias, c' . This conclusion also applied to the relationship between the two accuracy indices, ξ and d' .

Example 1. PISA Data on Mathematical Concepts

We were fortunate to have access to mathematical knowledge-claiming data from the Programme for International Student Assessment (PISA) questionnaire in 2012 (OECD, 2014), the most highly recognized cross-national study of student achievement. Students were asked to identify their familiarity with 16 mathematical concepts using a five-point scale, where 0 = *never heard of it*, 1 = *heard of it once or twice*, 2 = *heard of it a few times*, 3 = *heard of it often*, and 4 = *know it well and understand the concept*. Three specific yet nonexistent mathematical concepts (*proper number*, *subjunctive scaling*, and *declarative fraction*) were designed as foils. These same data proved useful in a recent study showing that accuracy and bias levels varied substantially across cultures (Fell et al., 2019).

We selected the Hong Kong subsample of 2936 students. The mean scores for the 13 real concepts ranged from 2.3 to 4.7 ($M = 3.6$), while those for the three foils ranged from 1.8 to 2.4 ($M = 2.0$). Since the SDT has been applied to analyze this dataset (OECD, 2014), we further

Table 2. Item slope estimates of knowledge bias and knowledge accuracy in Example 1.

Item	Mathematical Concept	λ_{i1}		λ_{i2}		$\left \frac{\lambda_{i1}}{\lambda_{i2}}\right $
		Estimate	SE	Estimate	SE	
ST62Q01	Exponential function	0.772	0.022	0.608	0.022	1.27
ST62Q02	Divisor	0.393	0.042	1.859	0.042	0.21
ST62Q03	Quadratic function	0.849	0.023	0.459	0.023	1.85
ST62Q04	Proper number*	1.513	0.029	0.335	0.029	4.52
ST62Q06	Linear equation	0.778	0.023	0.318	0.023	2.44
ST62Q07	Vectors	0.772	0.022	0.105	0.022	7.35
ST62Q08	Complex number	0.639	0.020	0.299	0.020	2.14
ST62Q09	Rational number	1.212	0.029	1.677	0.029	0.72
ST62Q10	Radicals	0.948	0.025	1.324	0.025	0.72
ST62Q11	Subjunctive scaling*	0.597	0.020	0.015	0.020	41.17
ST62Q12	Polygon	0.558	0.035	1.595	0.035	0.35
ST62Q13	Declarative fraction*	1.330	0.027	0.422	0.027	3.15
ST62Q15	Congruent figure	0.385	0.018	0.479	0.018	0.80
ST62Q16	Cosine	0.624	0.020	0.151	0.020	4.14
ST62Q17	Arithmetic mean	0.423	0.017	0.356	0.017	1.19
ST62Q19	Probability	0.637	0.021	0.647	0.021	0.98

Note. * = intended foils. λ_{i1} = item slope on dimension “knowledge bias”, λ_{i2} = item slope on dimension “knowledge accuracy”.

implemented the MMOC and examined the consistency of the two techniques in reporting the two indices of knowledge accuracy and knowledge bias. The correlation between the d' and ξ estimates was .83, and that between the c' and θ estimates was .95. In short, the two approaches tapped the same constructs.

As noted earlier, reliabilities are not available for the SDT indices. Using the MMOC, however, we obtained EAP reliability² estimates of .73 and .84 for accuracy and bias, respectively. Table 2 shows the estimates of the slope (λ_{i1} and λ_{i2}) parameters for each of the 16 items under the MMOC. Thus, the bias saturation ratio of $\lambda_{i1}/\lambda_{i2}$ was available for every single mathematical concept. Two of the real concepts, *divisor* and *polygon*, had the smallest ratios of $0.393/1.859 = 0.21$ and $0.558/1.595 = 0.35$, respectively, suggesting that their bias effects were relatively small. By contrast, the three foils (*proper number*, *subjunctive scaling*, and *declarative fraction*) displayed larger ratios (4.52, 41.17, and 3.15, respectively), suggesting substantial bias effects. The ratios for *vector* and *cosine*, meanwhile, were 7.35 and 4.14, respectively, suggesting that these items exhibited notable bias effects despite being real concepts. In brief, the results echoed Muszyński et al.’s (2021) findings that students’ responses to a real or fabricated mathematical concept reflected their knowledge accuracy and knowledge bias.

Although no statistical tests are currently available, the guidelines are straightforward. Items with low saturation ratios (well below 1.0) are the best future candidates for reals; items with high saturation rates (well above 1.0) make the best future foils. These guidelines hold regardless of whether the items were originally categorized properly. Rather than relabeling vector and cosine as foils, for example, we recommend dropping these two items from future PISA item sets. In sum, the MMOC results should prove useful in adjusting the overclaiming datasets after the fact.

Conclusion and Discussion

The overclaiming technique (OCT) was developed to assess individual differences in both the accuracy and exaggeration of knowledge (for a review, see [Paulhus, 2011](#)). Respondents indicate their level of familiarity with a set of items, some of which are real and some of which are foils. For each respondent, the endorsement rates of reals and foils are analyzed with formulas from SDT. Such formulas yield measures of individual differences in knowledge accuracy and knowledge bias. No item features are generated.

Comparison with the Williams–Zumbo Two-Stage Model

In an extension designed to address item features, [Williams and Zumbo \(2003\)](#) noted similarities between SDT and item response theory (IRT). They proposed a two-stage method and found it feasible to describe item functioning for knowledge accuracy, as calculated from overclaiming data. We also recognized the value of modeling item features for overclaiming data and proposed a new technique that exploits MIRT. Our proposed multidimensional model for overclaiming (MMOC) quantifies knowledge accuracy and knowledge bias as two independent latent traits.

Our model differs from the two-stage method in several ways. First, the latent trait variables in the two-stage method are the estimated factor scores of the factor analysis on pseudo items³, whereas the variables in the MMOC are estimated using raw items under the MIRT framework. Second, the item parameters of the two-stage method are obtained indirectly from a scatterplot showing the relationship between the factor score and a pseudo item. These parameters can be directly estimated using software (e.g., TAM). Third, the two-stage model adopted a nonparametric approach to describe item characteristics; by contrast, the MMOC is parametric (see [Equations \(3\)–\(6\)](#)).

Although both methods enable researchers to analyze overclaiming data in an IRT framework, the MIRT approach has distinct advantages over the two-stage approach. First, the MMOC yields separate scores for knowledge accuracy and bias, whereas the two-stage approach neglects the bias component. That neglect is unfortunate, considering that both accuracy and bias are essential factors in responding to OCQ items. In addition, fitting the MMOC is straightforward but offers a more in-depth analysis of psychometric properties. Third, the item parameter estimates obtained from the MMOC model can inform the winnowing or selection of real and foil items from a novel pool. For example, researchers can select more difficult items when a more capable group of respondents is to be evaluated.

Comparison with the SDT Approach

The results of the simulation study indicated that the corresponding person measures of SDT and the MMOC were highly intercorrelated, suggesting that these two approaches tap the same constructs. The key advantage of the MMOC is the provision of item and person features. From the item features, we derived a novel index labeled “bias saturation.” When applied to empirical studies (such as our two examples), the index provides a straightforward criterion for selecting the best reals (low saturation) and the best foils (high saturation) for follow-up studies.

The MMOC also elaborates on person features by providing measurement errors for the accuracy and bias indices. For example, the test reliabilities for these indices can be exploited to deattenuated correlation coefficients with outcome variables. In short, to achieve more accurate inferences, we recommend using MMOC person estimates to investigate associations with other variables (e.g., SDR scales and psychophysiological measurements) in future research.

One could argue that the arbitrary constraint of orthogonality between knowledge accuracy and knowledge bias in the MMOC is unrealistic. Holman and Glas (2005) provided an alternative constraint by fixing $\lambda_{i1} = 0$ for some items and $\lambda_{i2} = 0$ for others. In other words, the orthogonal constraint could be relaxed when at least one item purely measuring accuracy and another one purely measuring bias can be successfully identified; however, such items are impossible in the OCT context.

In principle, overclaiming data are no different from other rating-scale data. Because IRT has become the dominant psychometric approach for analyzing rating-scale data, other useful statistics are now available for practitioners to check the quality of measurement through IRT modeling. For example, item/person fit statistics that have been developed for categorical data in the IRT framework, such as $S-X^2$ (Orlando & Thissen, 2000; 2003) and \hat{P}_{non} (Douglas & Cohen, 2001), can be directly applied to overclaiming datasets. IRT analyses also provide insight into the invariance of item functioning across different samples (e.g., Wang, 2008). As we demonstrated, these tools can be used for overclaiming item development and evaluation.

Limitations and Alternatives

To date, a number of studies have supported the use of OCT as a practical and cost-effective approach for detecting SDR in academic, cross-cultural, and organizational research (for a review, see Paulhus, 2011). Nonetheless, this method has not always proven successful in personnel selection (Ludeke & Makransky, 2016). It now appears clear that a necessary condition for successful application is the development of effective foils (see Dubois & Paulhus, 2020). In the most thorough investigation conducted to date, Dunlop et al. (2019) varied the content of the overclaiming items. They showed that the bias index only predicts faking if (a) the stakes are high, and (b) items match the nature of the job application. For example, applicants who tend to exaggerate their musical knowledge are unlikely to exaggerate their chemistry knowledge when applying for a pharmacy job.

Nonetheless, other interpretations of overclaiming indices have been offered. Instead of impression management, the bias index has been interpreted as an exaggerated feeling of knowing (Atir et al., 2015; Clariana et al., 2016); that is, high scores reflect a subjective sense that many stimuli are familiar. Most compelling is Study 3 from Atir and colleagues, where expertise was experimentally manipulated and higher familiarity ratings ensued. Closely related interpretations of overclaiming bias include a high openness to experience (Dunlop et al., 2017) and uncritical credulousness (Pennycook & Rand, 2018).

The current article opens up other possible avenues for overclaiming research. The broadest question concerns the degree to which accuracy and bias are trait-like qualities. Temporal stability remains untested. Level of education was found to have little effect on overclaiming tendencies (Clariana et al., 2016). Our second empirical example also showed a negligible difference in overclaiming bias between 14-year-olds and college undergraduates. More convincing would be a within-subjects design to determine whether the rank order of person qualities is stable. Cross-situational stability is also a key issue: Dunlop et al. (2020) began the task with person features, but the MMOC raises the question of stability in item features.

Some 20 years ago, the cross-topic consistency of accuracy and bias indices was demonstrated using the 10 topics of the OCQ-150 (see Paulhus et al., 2003); however, those topics were all academic in nature.⁴ Investigation of more diverse content domains may conclude that a single pair of θ and ξ values is insufficient to capture an individual's person features. The bias saturation of items will likely vary across content domains. Moreover, neglected in the current analysis are advances in rating-response styles (e.g., Bolt & Johnson, 2009; Jonas & Markon, 2019). Already, more general MIRT models have been developed to accommodate the OCT paradigm and various

response styles (e.g., Chen & Jin, 2022). In addition, inattentive responding (e.g., Jin et al., 2018) can also be integrated into IRT analyses of overclaiming. Finally, other software packages warrant evaluation to compare their performances and efficiencies in conducting such analyses. Notwithstanding these issues, we believe that the MMOC offers a substantial advance in the analysis of overclaiming data.

Acknowledgments

The authors would like to thank Dr Liu Li for granting permission to use the Chinese idiom data.

Author Note

Visiting Scholar, The Measurement, Evaluation, and Research Methodology (MERM) program University of British Columbia, Canada.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Delroy L. Paulhus was supported by Social Sciences and Humanities Research Council of Canada (Grant 435–2015-0417). Ching-Lin Shih was supported by Ministry of Science and Technology of Taiwan (grant number MOST 108-2918-I-110 -005 and MOST 110–2410-H-110–044).

ORCID iDs

Kuan-Yu Jin  <https://orcid.org/0000-0002-0327-7529>

Ching-Lin Shih  <https://orcid.org/0000-0002-0655-849X>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. This negation makes more positive scores represent greater bias (i.e., knowledge exaggeration).
2. According to the TAM manual, the EAP reliability is defined as $\nu/(s + \nu)$, where ν and s denote the variance of the estimates and average of the standard errors, respectively.
3. Recall that these pseudo items are parcels of items that include both reals and foils.
4. Content extensions include vocabulary (Ziegler et al., 2013), and religious authorities (Jones et al., 2020).

References

- Ackerman, P. L., & Ellingsen, V. J. (2014). Vocabulary overclaiming — a complete approach: Ability, personality, self-concept correlates, and gender differences. *Intelligence*, 46(1), 216–227. <https://doi.org/10.1016/j.intell.2014.07.003>
- Adams, R. J., Wilson, M., & Wang, W.-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1–23. <http://doi.org/10.1177/0146621697211001>

- Atir, S., Rosenzweig, E., & Dunning, D. (2015). When knowledge knows no bounds: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, 26(8), 1295–1303. <https://doi.org/10.1177/0956797615588195>
- Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, 116(1), 148–162. <http://doi.org/10.1016/j.obhdp.2011.05.006>
- Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, 33(5), 335–352. <https://doi.org/10.1177/0146621608329891>
- Chen, H.-F., & Jin, K.-Y. (2022). The impact of item feature and response preference in a mixed-format design. *Multivariate Behavioral Research*, 57(2–3), 208–222. <https://doi.org/10.1080/00273171.2020.1820308>
- Clariana, M., Castelló, A., & Cladellas, R. (2016). Feeling of knowing and over-claiming in students from secondary school to university. *Learning and Individual Differences*, 49, 421–427. <https://doi.org/10.1016/j.lindif.2016.05.008>
- Cronbach, L. J. (1949). *Essentials of psychological testing*. Harper.
- DeCarlo, L. T. (2020). An item response model for true–false exams based on signal detection theory. *Applied Psychological Measurement*, 44(3), 234–248. <https://doi.org/10.1177/0146621619843823>
- DeCarlo, L. T. (2021). A signal detection model for multiple-choice exams. *Applied Psychological Measurement*, 45(6), 423–440. <https://doi.org/10.1177/01466216211014599>
- Douglas, J., & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, 25(3), 234–243. <https://doi.org/10.1177/01466210122032046>
- Dubois, P., & Paulhus, D. L. (2020). *The VOCE: A measure of vocabulary over-claiming in English* [manuscript under review]. University of British Columbia.
- Dunlop, P. D., Bourdage, J. S., de Vries, R. E., Hilbig, B. E., Zettler, I., & Ludeke, S. G. (2017). Openness to (reporting) experiences that one never had: Overclaiming as an outcome of the knowledge accumulated through a proclivity for cognitive and aesthetic exploration. *Journal of Personality and Social Psychology*, 113(5), 810–834. <https://doi.org/10.1037/pspp0000110>
- Dunlop, P. D., Bourdage, J. S., de Vries, R. E., McNeill, I. M., Jorritsma, K., Orchard, M., Austen, T., Baines, T., & Choe, W.-K. (2020). Liar! Liar! (when stakes are higher): Understanding how the overclaiming technique can be used to measure faking in personnel selection. *Journal of Applied Psychology*, 105(8), 784–799. <http://doi.org/10.1037/apl0000463>
- Fell, C. B., König, C. J., Jung, S., Sorg, D., & Ziegler, M. (2019). Are country level prevalences of rule violations associated with knowledge overclaiming among students? *International Journal of Psychology*, 54(1), 17–22. <https://doi.org/10.1002/ijop.12441>
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Sage.
- Furr, R. M., & Funder, D. C. (2007). Behavioral observation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 273–291). Guilford Press.
- Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing competing claims about overclaiming. *Intelligence*, 81, 101470. <https://doi.org/10.1016/j.intell.2020.101470>
- Holden, R. R., & Kroner, D. G. (1992). The relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment*, 4(2), 170–173. <http://doi.org/10.1037/1040-3590.4.2.170>
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63(2), 272–279. <http://doi.org/10.1037/0022-3514.63.2.272>

- Holman, R., & Glas, C. A. W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *The British Journal of Mathematical and Statistical Psychology*, 58(1), 1–17. <http://doi.org/10.1111/j.2044-8317.2005.tb00312.x>
- Hülür, G., Wilhelm, O., & Schipolowski, S. (2011). Prediction of self-reported knowledge with overclaiming, fluid and crystallized intelligence and typical intellectual engagement. *Learning and Individual Differences*, 21(6), 742–746. <https://doi.org/10.1016/j.lindif.2011.09.006>
- Jeon, M., & De Boeck, P. (2016). A generalized item response tree model for psychological assessments. *Behavior Research Methods*, 48(3), 1070–1085. <https://doi.org/10.3758/s13428-015-0631-y>
- Jin, K.-Y., Chen, H.-F., & Wang, W.-C. (2018). Mixture item response models for inattentive responding behavior. *Organizational Research Methods*, 21(1), 197–225. <https://doi.org/10.1177/1094428117725792>
- Jin, K.-Y., & Wang, W.-C. (2014). Generalized IRT models for extreme response style. *Educational and Psychological Measurement*, 74(1), 116–138. <http://doi.org/10.1177/0013164413498876>
- Jonas, K. G., & Markon, K. E. (2019). Modeling response style using vignettes and person-specific item response theory. *Applied Psychological Measurement*, 43(1), 3–17. <https://doi.org/10.1177/0146621618798663>
- Jones, D. N., Neria, A. L., Helm, F. A., Sahlan, R. N., & Carré, J. R. (2020). Religious overclaiming and support for religious aggression. *Social Psychological and Personality Science*, 11(7), 1011–1021. <https://doi.org/10.1177/1948550620912880>
- Kam, C., Risavy, S. D., & Perunovic, W. E. (2015). Using over-claiming technique to probe social desirability ratings of personality items: A validity examination. *Personality and Individual Differences*, 74, 177–181. <https://doi.org/10.1016/j.paid.2014.10.017>
- Kiefer, T., Robitzsch, A., & Wu, M. (2020). *TAM: Test Analysis modules* (Version 3.5-19) [Computer software]. <https://cran.r-project.org/web/packages/TAM/index.html>
- Liu, L., & Cheung, H. T. (2018). The partial productivity of schematic idioms in Chinese: Evidence from speakers' judgment. *Leukemia & Lymphoma*, 19(1), 80–116. <http://doi.org/10.1075/lali.00004>
- Ludeke, S. G., & Makransky, G. (2016). Does the over-claiming questionnaire measure overclaiming? Absent convergent validity in a large community sample. *Psychological Assessment*, 28(6), 765–774. <http://doi.org/10.1037/pas0000211>
- MacMillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Erlbaum.
- Müller, S., & Moshagen, M. (2018). Overclaiming shares processes with the hindsight bias. *Personality and Individual Differences*, 134, 298–300. <https://doi.org/10.1016/j.paid.2018.06.035>
- Müller, S., & Moshagen, M. (2019). True virtue, self-presentation, or both?: A behavioral test of impression management and overclaiming. *Psychological Assessment*, 31(2), 181–191. <https://doi.org/10.1037/pas0000657>
- Muszyński, M. (2020). *Validity of the overclaiming technique as a method to account for response bias in self-assessment questions: Analysis on the basis of the PISA 2012 data* (Unpublished doctoral thesis). Jagiellonian University. https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/251494/muszynski_validity_of_the_overclaiming_technique_as_a_method_2020.pdf?sequence=1&isAllowed=y
- Muszyński, M., Pokropek, A., & Żółtak, T. (2021). Structural validity of overclaiming scores: Analysing PISA 2012 data. *Psychological Test and Assessment Modeling*, 63(1), 119–145. https://www.psychologie-aktuell.com/fileadmin/download/ptam/1-2021/Seiten_aus_PTAM_2021-1_ebook_6.pdf
- OECD. (2014). *PISA 2012 technical Report*. OECD.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24(1), 50–64. <https://doi.org/10.1177/01466216000241003>
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27(4), 289–298. <https://doi.org/10.1177/0146621603027004004>
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press.

- Paulhus, D. L. (2011). Overclaiming on personality questionnaires. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151–164). Oxford University Press.
- Paulhus, D. L. (2017). Socially desirable responding on self-reports. In V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences*. Springer. http://doi.org/10.1007/978-3-319-28099-8_1349-1
- Paulhus, D. L., & Dubois, P. J. (2014). Application of the overclaiming technique to scholastic assessment. *Educational and Psychological Measurement*, 74(6), 975–990. <https://doi.org/10.1177/0013164414536184>
- Paulhus, D. L., & Harms, P. D. (2004). Measuring cognitive ability with the overclaiming technique. *Intelligence*, 32(3), 297–314. <http://doi.org/10.1016/j.intell.2004.02.001>
- Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, 84(4), 890–904. <http://doi.org/10.1037/0022-3514.84.4.890>
- Paulhus, D. L., & Holden, R. R. (2010). Measuring self-enhancement: From self-report to concrete behavior. In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 227–246). Oxford University Press.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). Guilford Press.
- Pennycook, G., & Rand, D. G. (2018). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2), 185–200. <https://doi.org/10.1111/jopy.12476>
- Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81(2), 93–103. https://doi.org/10.1207/S15327752JPA8102_01
- Sedikides, C., & Alicke, M. D. (2019). The five pillars of self-enhancement and self-protection. In R. M. Ryan (Ed.), *Oxford handbook of human motivation* (pp.307–319). Guilford Press.
- Sorkin, R. D. (1999). Spreadsheet signal detection. *Behavior Research Methods, Instruments, and Computers*, 31(1), 46–54. <https://doi.org/10.3758/BF03207691>
- Swami, V., Papanicolaou, A., & Furnham, A. (2011). Examining mental health literacy and its correlates using the overclaiming technique. *British Journal of Psychology*, 102(3), 662–675. <https://doi.org/10.1111/j.2044-8295.2011.02036.x>
- Swets, J. A. (1964). *Signal detection and recognition by human observers*. Wiley.
- Vonkova, H., Papajoanu, O., & Stipek, J. (2018). Enhancing the cross-cultural comparability of self-reports using the overclaiming technique: An analysis of accuracy and exaggeration in 64 cultures. *Journal of Cross-Cultural Psychology*, 49(8), 1247–1268. <https://doi.org/10.1177/002202211878704>
- Wang, W.-C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement*, 9(4), 387–408.
- Wang, W.-C., Chen, H.-F., & Jin, K.-Y. (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157–178. <https://doi.org/10.1177/0013164414528209>
- Wang, W.-C., Wilson, M., & Shih, C.-L. (2006). Modeling randomness in judging rating scales with a random-effects rating scale model. *Journal of Educational Measurement*, 43(4), 335–353. <http://doi.org/10.1111/j.1745-3984.2006.00020.x>
- Williams, K. M., & Zumbo, B. D. (2003). Item characteristic curve estimation of signal detection theory-based personality data: A two-stage approach to item response modeling. *International Journal of Testing*, 3(2), 189–213. http://doi.org/10.1207/S15327574IJT0302_7
- Ziegler, M., Kemper, C., & Rammstedt, B. (2013). The vocabulary and overclaiming test (VOC-T). *Journal of Individual Differences*, 34(1), 32–40. <https://doi.org/10.1027/1614-0001/a000093>
- Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). *New perspectives on faking in personality assessment*. Oxford University Press.