2024 ■ Vol. 20 ■ no. 1

Evaluating assessment via item response theory utilizing information function with R

Teck Kiang Tan [a]

[a] Centre for Embodied Learning and Living, National University of Singapore

Abstract ■ Item and test information functions that measure the reliability of an assessment via Item response theory (IRT) are described in the paper for the practitioners. While the four parameters binary IRT models are frequently used, their relationship to the precision level is not commonly discussed. More notably, the benefits, limitations, and constraints of the information approach have not been fully examined systematically. On this basis, with useful and practical examples, the paper formally introduces the graphical approach of presenting item and test information functions that could be easily carried out using the irt R package. The simple R syntax is illuminated throughout the text to show the separate item parameter effects and the combinational and offsetting effects on the information when all the item parameters are used in an assessment. In particular, the characteristics of the 4PL information function that have not been paid much attention to are highlighted and illustrated about its functionality and application. The paper ends with a guide on the information approach to the selection of the items and setting up an assessment. The scope, limitations, and constraints of the graphical approach are also discussed.

Keywords ■ Test information function, Item information function, Item response theory. **Tools** ■ R.

alsttk@nus.edu.sg

10.20982/tqmp.20.1.p033

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ One anonymous reviewer

Introduction

Item Response Theory (IRT) is an analytical method for analyzing, constructing, and evaluating assessments, tests, psychological inventories, and any other types of inventories. It is also named latent trait analysis and modern test theory, which refers to a family of models that attempt to explain the relationship between latent traits and their manifestation, the test items. When administering and setting up an assessment using IRT models, the precision of the assessment is core to determining the reliability of the assessment (Baker & Kim, 2017; Hambleton & Swaminathan, 2013). However, the common practice of validation and evaluation of assessment generally reports the information of an assessment but does not describe how the information is derived to attain an acceptable level of precision. While the relationship among the IRT models, the reliability, and the information of an assessment are crucial, they are seldom discussed and stressed in detail to associate them to evaluate and set up an assessment.

Since the use of Computer Adaptive Testing (CAT) application has become common in assessment (e.g. Yao, 2019), examining the relationship between the parameters of the four IRT binary models and information has also become central. The inclusion of the inattention parameter into practical applications especially in CAT has directed the conception of information in a more relevant context and becomes more valuable and accurate in determining the precision of an assessment (Culpepper, 2017; Kalkan, 2022; Rulison & Loken, 2009; Zheng et al., 2021). Hambleton and Swaminathan (2013) describe the association between information and its application in the precision of measurement however stop at the three-parameter Logistic (3PL) model and do not extend to the four-parameter Logistic (4PL) model.

Mixing all four IRT models (One Parameter Logistic (1PL), Two Parameter Logistic (2PL), Three Parameter Logistic (3PL), and Four Parameter Logistic (4PL)) into an assessment has become more common in practice. On this basis, this combinational approach of including item param-

Table 1 ■ Item Information Function of IRT Models

IRT Model	Item Information Function $I_i(\theta)$
1PL	$D^2 p_i(\theta)\, q_i(\theta)$
2PL	$D^2 a_i^2 p_i(\theta) q_i(\theta)$
3PL	$\frac{D^2 a_i^2 q_i(\theta)}{p_i(\theta)} \left\{ \frac{(p_i(\theta)-c_i)^2}{(1-c_i)^2} \right\}$
4PL	$D^2 a_i^2 \frac{(p_i(\theta)-c_i)^2}{(d_i-c_i)^2} \frac{(d_i-p_i(\theta))^2}{p_i(\theta)q_i(\theta)}$

eters of the four IRT models in an assessment requires the understanding of first, the relationship of these four item parameters and information. Second, the combinational and offsetting effects of inclusion of all the four-parameter items into an assessment. The main reason is that when a combination takes place, it could result of an opposite effect on the information. As such, the way the combination takes place needs to be understood on their effects of the information magnitude. Third, examining information in an assessment can be time-consuming if there are no straightforward software tools with simple syntax to determine and examine the level of information. Fourth, software that is able and easily carries out sensitivity analyses to examine the effects more flexibly becomes a viable practical choice. Fifth, adapting the graphical approach recommended by Hambleton and Swaminathan (2013) is helpful for educators and assessors to examine and set up an assessment by visualizing the information effect.

The main goals of the paper are thus (a) to show the association of item and test information to the parameters of the four IRT models; (b) to highlight the magnitudes of the combinational and countering effects of using all four IRT models in an assessment; (c) to show the R syntax to generate the test and information functions; (d) to use the graphical approach to generating the item and test information curves, and (e) to provide practical strategic guide for the item selection into an item bank.

Item Response Theory Model and Item Information Function

In the IRT context, precision is generally referred to as the information of an item or a test, for the former is about the precision level of an item, and the latter of an assessment. The term information in IRT is defined as the reciprocal of precision, the variability around the value of the parameter estimate, stated in Equation (1) where $I(\theta)$ refers to the item information function given the ability θ and SE refers to the standard error. Here, SE is specified as the level of uncertainty of the estimate that directly measures the information of a test item, given the value of θ. High information about an item represents more precision of the item and vice versa.

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}} \tag{1}$$

As information is defined, it is not a point estimate but varies over a region by the given θ value. This way of specifying information has an obvious advantage as precision is not uniform across the entire range of the ability θ for scores at the edges of the test's range generally have more error than scores closer to the middle of the estimate which normally peaks at a theta position. IRT not only advances the concept of item and test information to replace the classical way of defining reliability, but the information is also a function of the model parameters. Mathematically, information is the first derivative of the item characteristic curve (DeMars, 2018). Table 1 lists the item information function of the one-parameter logistic (1PL; Rasch, 1960), the two-parameter logistic (2PL; Birnbaum, 1958, 1968), the three-parameter logistic (3PL; Birnbaum, 1968) model, and the four-parameter logistic (4PL; Barton & Lord, 1981) model in which $I_i(\theta)$ represents the item formation function of item i, $p_i(\theta)$ represents the probability of getting a correct answer given the value of θ, $p_i(\theta) = 1 - q_i(\theta)$, D is a scaling factor whose value is generally set to 1.7, a represents the discrimination parameter, b represents the difficulty parameter, c represents the guessing parameter, and d represents the inattention parameter (de Ayala, 2009; Hambleton & Swaminathan, 2013; SAS/STAT, 2023).

The general procedure for understanding the relationship of ability, parameter estimates, and information is best to plot the item information function on the y-axis and the ability on the x-axis for an item to examine the level of information over the ability scale range. Here, it is referred to as the graphical approach to examining information function.

Item Information Function – 1PL

The item information function for 1PL is the product of the probability of getting a correct and incorrect response $[p_i(\theta) \times q_i(\theta)]$. As such, the item information peaks when $p_i(\theta) = q_i(\theta) = 0.5$, with $I_i(\theta) = 0.25D^2$ (Table 1). The theta location of this peak, θ_{max}, for 1PL is at $\theta_{max} = b_j$, that is, the estimated value of the difficulty parameter is where the theta location of the information peaks. Figure

Figure 1 ■ Item Information function - 1PL Model

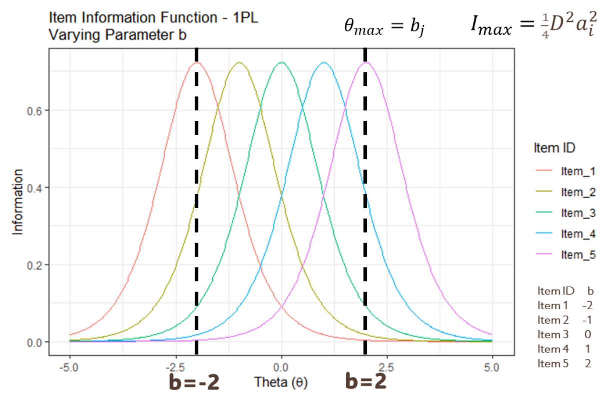
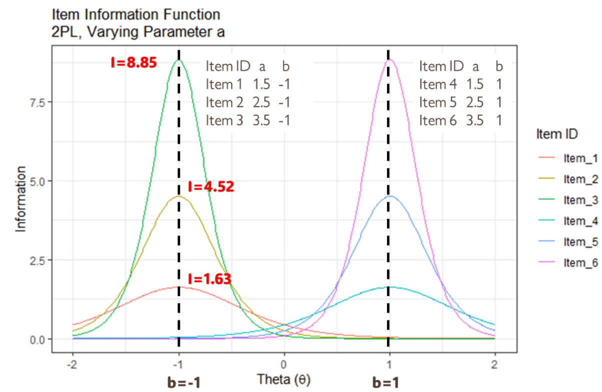


Figure 2 ■ Item Information Function – 2PL, The Effect of Discrimination Parameter



1 shows the maximum values of five items where the peak of the item information function occurs when $\theta = b$. For instance, when $b = -2$, the information of item 1 peaks at $\theta = -2$, and for $b = 2$, the information of item 5 peaks at $\theta = 2$. These five information functions in Figure 1 show the peak happens at the b logit theta. The value of information reduces from the peak and becomes smaller when it is further away from the b logit theta.

The R syntax is rather straightforward to generate the item functions of these 5 items with the difficulty parameters that range from -2 to 2 with an incremental of 1 logit. See the listing 1 in the appendix. First, the R package `irt` (Gonulates, 2022) is loaded using the library function. Under the function `itempool`, there are two arguments required to specify. The first parameter, `b=` specifies the values of the difficulty parameter, that runs from -2 to 2 with an incremental value of 1. The `seq` function specifies this running numeric number. The second parameter, `D = 1.7`, specifies the scaling value. These five difficulty parameters are stored in the R object called Item01. The `plot_info` function reads in the object `Item01` to produce the five information curves graphed in Figure 1. The argument `title` prints the title of the plot on top of the graph.

Item Information Function – 2PL

For 2PL, the discrimination parameter plays a vital role in determining the information of an item. As this parameter is specified as the square of the discrimination parameter, a_i^2, in the information formula, it indicates the larger the discrimination parameter the greater the information, and the information grows much higher when this parameter gets larger. Figure 2 graphs 6 item functions at two theta locations of -1 and 1 logit. As both sets of difficulty parameters contain the same values, the shapes of the information are the same except for their theta locations. When the value of $a = 1.5$, the maximum value of information is 1.63. The maximum information value grows to 4.52 when $a = 2.5$ and further increases to 8.85 when $a = 3.5$. This results in an increment of a unit of the discrimination parameter that almost doubles the value of the maximum value of information. The general conclusion is that the higher the value of the discrimination parameter, the larger the maximum value of information. Figure 2 also indicates the higher the discrimination parameter value, the narrower the spread of information. It also indicates where the information peaks at p=q=0.5, this location is equal to the difficulty parameter value, similar to that of 1PL.

The syntax of generating item information of the 2PL is similar to that of the 1PL (see appendix, Listing 2). The additional specification is to add in the coefficients of the discrimination parameter after the argument `a=`. The argument `theta_range` is added in the function `plot_info` to specify the range of theta to print the item information function. The `rep` function repeats the value specified in the first argument and the second argument states the number of repeats. For instance, `rep(-1,3)` generates a vector of three values of -1. Similarly, the function `rep(1,3)` generates a vector consisting of three values of 1.

Sometimes, identifying the value of the information is useful for understanding their relationship to the theta ability. The listing 3 in the appendix provides the syntax to generate the information functions for the 6 items specified in the object Item02. The object `thetaRange` specifies the information to print by restricting them to 5 points of theta (-2, -1, 0, 1, and 2). The function info generates the information values and stores them in an object named

Figure 3 ■ The Effect of Guessing Parameter on Information – 3PL

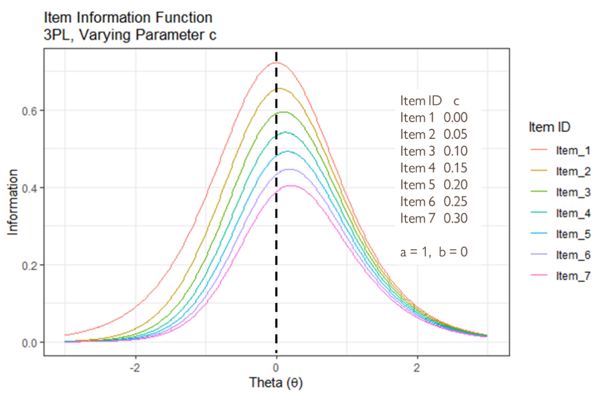
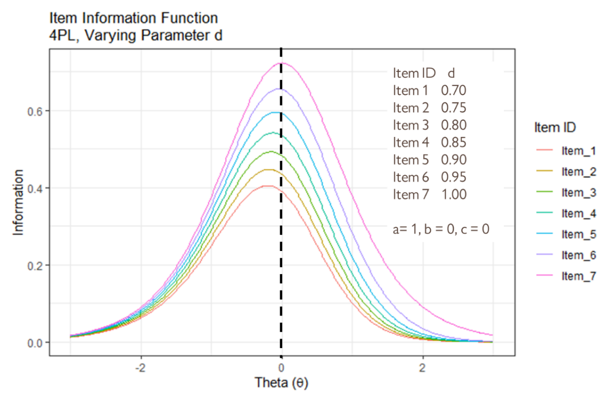


Figure 4 ■ The Effect of Inattention Parameter on Information – 4PL



`ItemInfo02`. The rownames function gives the names for each row of the object `ItemInfo02`. The round function further restricted four decimals to be printed. The output of the `ItemInfo02` (Appendix, Listing 3) prints the maximum information values for the 6 items that correspond to the values of information printed in Figure 2. Those highlighted in red are the maximum information values for the six items.

Item Information Function – 3PL

Concerning the impact of the guessing parameter on precision, Figure 3 shows that the information becomes lower for a higher value of the guessing parameter. For instance, the highest information curve in Figure 3 is with values of `c=0` (Item 1). The information reduces to its lowest for `c=0.3` (Item 7). It is noted the theta location where the information peak is no longer fixed to the value of the difficulty parameter where `b=0`. The dotted vertical line in Figure 3 indicates the location of the `theta=0` but the maximum value of information falls away from this position. When the value of the guessing parameter increases, the information curve shifts toward the right. In conclusion, while the maximum information for 1PL and 2PL occurs at b_i, 3PL occurs at

$$b_i + \frac{1}{Da_i}\left\{ln\frac{1+\sqrt{1+8c_i}}{2}\right\}$$

(Hambleton & Swaminathan, 2013), indicating the larger the value of the guessing parameter, the larger the value of θ_{max} away from the value of the difficulty parameter by the amount of

$$\frac{1}{Da_i}\left\{ln\frac{1+\sqrt{1+8c_i}}{2}\right\}.$$

The R syntax of Listing 4 in the Appendix which generates the pool of items in Figure 3 is similar to that of the 1PL and 2PL (Appendix, Listing 4) with additional specification of the `c=` argument.

Item Information Function – 4PL

The impact of the inattention parameter on precision is shown in Figure 4. The relationship between the inattention item parameter and information is that the lower the value of the inattention parameter, the lower the information. Same as the 3PL, the location of the maximum value of information is also not positioned at the value of the difficulty parameter. In contrast to 3PL, the peak is in the opposite direction concerning the value of the inattention parameter for the smaller the value of the inattention parameter, the lower the theta location from the difficulty parameter value. In short, the peak of the information for the 4PL shifted to the left while the 3PL shifted in the opposite direction to the right. In summary, the lower the value of the inattention parameter, the lower the information, and the peak location shifted away from the difficulty parameter to the left at the lower theta location.

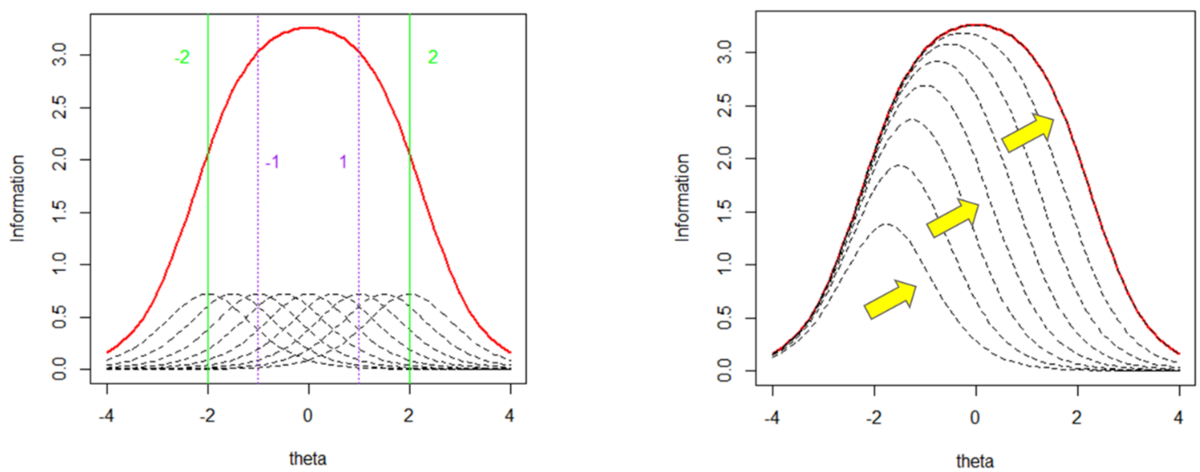
The syntax for the 4PL to generate Figure 4 is given in Listing 5 in the Appendix by adding in the `d=` argument to specify the inattention parameter values.

In summary, the location and amount of information of an item largely depends on the values of the parameter for each parameter provides different precision about their position and magnitude. For 1PL, the information peaks at the ability level location equal to the difficulty parameter. For 2PL, the larger the value of the discrimination parameter, the greater the information, and the taller and narrower the information curve. For 3PL, the smaller the value of the guessing parameter, the greater the information. On the contrary, for 4PL, the larger the value of the inattention parameter, the greater the information.

Figure 5 ■ 1PL - Test and Cumulative Test Information

(a) b vary from -2 to 2 with increments of 0.5

(b) b vary from -2 to 2 with increments of 0.5



Test Information Function for Constructing An Assessment

Understanding the psychometric properties between the item parameter and item information function helps to determine the level of precision at the item level. However, the precision of an assessment is about the combination of items into a test. This leads to the aggregation of the item information to provide global precision evidence leading to the selection of items for an assessment. As such, the selection of a list of items for an assessment has to be carefully considered for the IRT model(s) used and their parameters to ensure the assessment is in line with the objectives of constructing an assessment (Hambleton & Swaminathan, 2013). Due to local independence, item information functions are additive. Thus, the test information function in the IRT context is simply by summing up the item information functions at each ability level (θ) stated in equation (2) below (Moghadamzadeh et al., 2011; Reise & Waller, 2009).

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta) \tag{2}$$

where $I(\theta)$ is the amount of test information at an ability θ level, $I_i(\theta)$ is the amount of information for item i at ability level θ, and N is the number of items in a test.

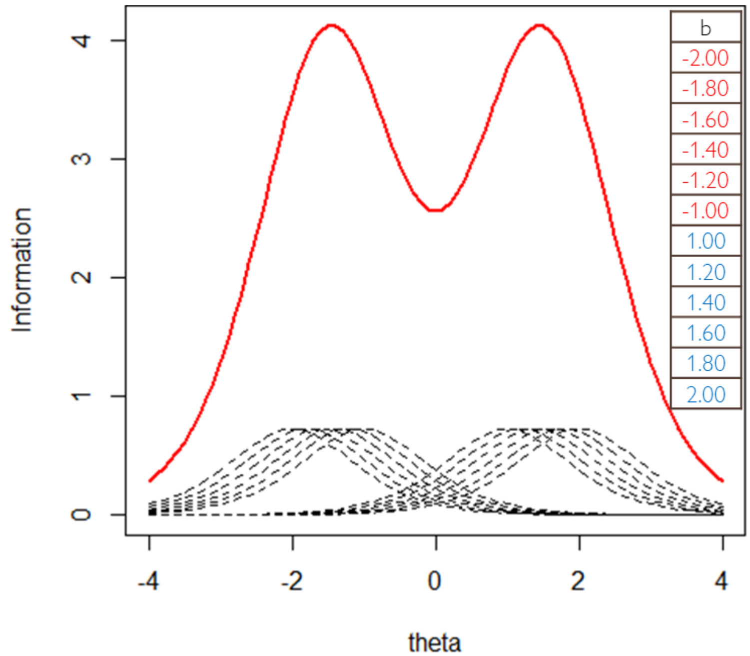
As a test information function is a summation of the item information functions, this property is useful as it facilitates the selection of items for inclusion in an assessment such that the precision of measurement for a test is maximized at the specified trait range. For instance, when choosing a group of students with lower ability traits and wishing to distinguish them more precisely, the test should reflect by including substantially more easy items around the relevant ability region. In short, the selection of items has to be in line with the purpose and consideration of their information to generate the assessment.

The following subsections illustrate seven examples to show how the test information function could be used to set up an assessment using the four binary IRT models. The first example shows the building up of a test information function using 1PL to form an assessment that consists of students with ability level that ranges from -2 to 2 logit. The item, test, and cumulative test information functions are graphed to show how the test information function is formed. The second example shows the formation of an assessment with two ability groups of examinees. The third example focuses on the selection of items for the high-ability examinees aiming to distinguish them more precisely without neglecting the low-ability examinees, and the fourth example concentrates on distinguishing the low-ability examinees which also include the high-ability examinees, based on 2PL. The fifth example shows the impact of the guessing parameter on the test information function. The sixth example shows the impact of the inattention parameter, and the seventh example focuses on the impact of varying both the guessing parameter and the inattention parameter. The last example illustrates the impact of increasing the number of items on the test information function.

Setting An Assessment using 1PL – Illustrating Cumulative Process to Form Test Infomation

Figure 5 graphs the information curves of a set of 9 items using 1PL with difficulty parameters ranging from -2 to 2

Figure 6 ■ Two Distinct Ability Groups – Item and Test Information Function



logits with an incremental of 0.5 logits of an assessment. The 9 blue dotted curves in the left panel of Figure 5 are the item information functions and the red curve on top of them is the test information function that sums up all the 9 information values of the item information functions. The resulting red-colored test information curve indicates the precision is best from -1 to 1 logit region enclosed by the two purple-colored dotted lines. High information could also be found from the -2 to 2 logits region enclosed by the two green-colored lines, and the two tail ends with lesser information reaching zero probability when they are closer to and below 4 and -4 logit.

The right panel of Figure 5 shows how the 9 item information functions are incrementally shaped to form the eventual test information function starting from the item information curve of the difficulty parameter specified as -2 logit, adding in the -1.5 logit item information, moving on to the -1 logit item information, and ending with the red-colored test information curve that includes the last item information curves with 2 logits. The arrows help to indicate the direction of expansion of the incremental cumulative movement. The conclusion is that the selection of the test items has to be in line with the expected examinee's ability such that the items selected are directly linked to the difficulty parameter when using the 1PL model. This example shows an assessment that aims to a group of examine's ability range from -2 to 2 logits, explained using both item and test information graphing, the graphical information approach.

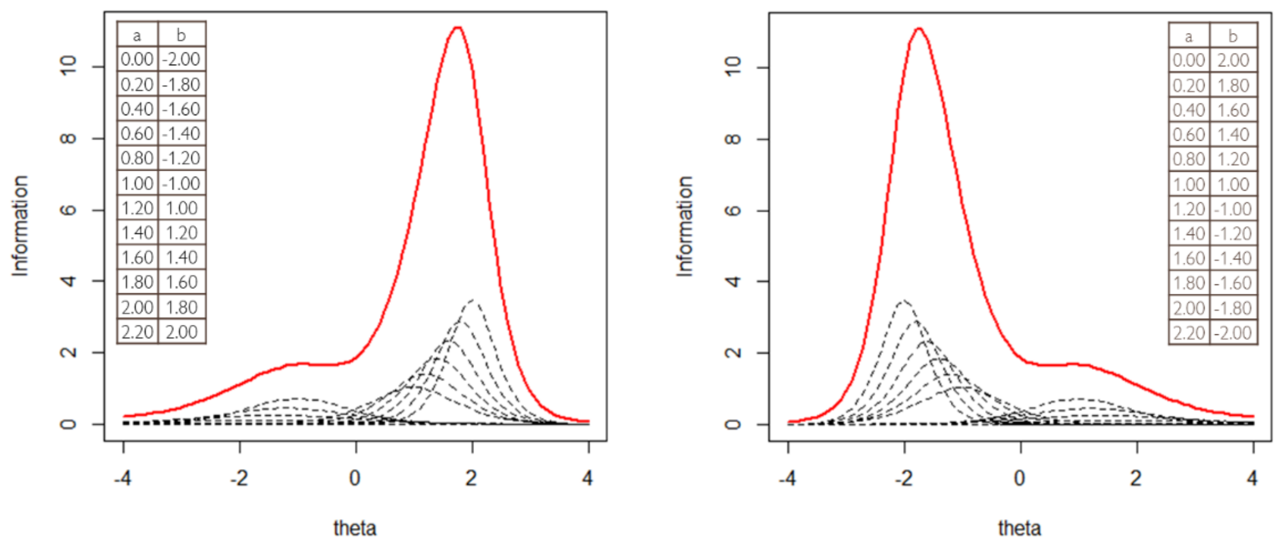
The R syntax to draw the test information function curve (given in Listing 6 in the Appendix) is to use the base R plot function by specifying the axis as theta and the y-axis as `Test1_Info` which stores the information of these 9 items. The ylim argument sets the range of the y-axis from zero and the maximum value of information using the max function. The first for loop function generates the item information function curves by looping it 9 times to generate the item test function, reading from the object `Test1_Info`. The second loop generates the cumulative test information functions using the rowSums function.

Focus on Two Distinct Abilities Examinee Groups

When facing two groups of examinees having distinct abilities, a direct solution is to select two clusters of items targeted to the two separate groups. Figure 6 graphs the test and item information functions of the two groups. The test information shows two peaks. The left peak is due to the inclusion of items with difficulty parameters ranging from -2 to -1 logits. The right peak is due to the inclusion of items with difficulty parameters ranging from 1 to 2 logits. This strategy ensures the precision of both lower and higher ability examines are taken into account for exami-

Figure 7 ■ Test & Item Information – 2PL

(a) Concentrate on Distinghising Low-Ability Examinees **(b)** Concentrate on Distinguishing High-Ability Examinees



nees with ability round -1.5 and +1.5 ability logit levels. The example illustrates item selection for an assessment when using the 1PL model has to match the difficulty level of the item to examinees who sit for the assessment.

Similar to the syntax of producing Figure 5, the syntax to produce Figure 6, given in Listing 7, is obtained by changing the values of the item parameters included in the `itempool` function. The R syntax to generate the item and test information functions is simply to include two sets of seq functions that cover the specified range of the difficulty parameters.

Concentrating on A Group of Examinees Without Neglect the others

Examinees who sit for an assessment may be quite diverse in their abilities. But if there is a focus on a particular group of examines to distinguish them more finely in their abilities without neglecting the rest, choosing items from an item bank with the consideration of both discrimination and difficulty parameters using the 2PL becomes strategic. For instance, to generate an assessment that aims to distinguish the high-ability examinees more precisely but also includes the lower-ability examinees. On the other hand, setting up an assessment that seeks to distinguish the low-ability examinees more thinly but also includes the high-ability examinees is also not uncommonly found. One straightforward tactical approach is to create a left-skewed test information curve for the former scenario and the latter with a test information curve that skews to the right.

The left and right panels of Figure 7 respectively depict these two situations using the 2PL model. The basic idea, for the former scenario, is to include a list of high discrimination and yet difficult items such that a small change in ability can be easily detected to distinguish the high-ability examines, and concurrently include a list of easy items for the lesser ability examinees. The left panel of Figure 7 shows this scenario with a left-skewed test information curve. The greatest benefit of adopting this strategy is that it allows the assessment not to be overloaded by including a large number of items in an assessment but to provide a list of high discrimination items according to the ability level of the group that needed to be thinly differentiated. On the contrary, the right panel of Figure 7 shows the opposite scenario with highly discriminating easy items to allow the distinction of the lower-ability examinees thinly and yet concurrently include the difficult items for the higher-ability examinees.

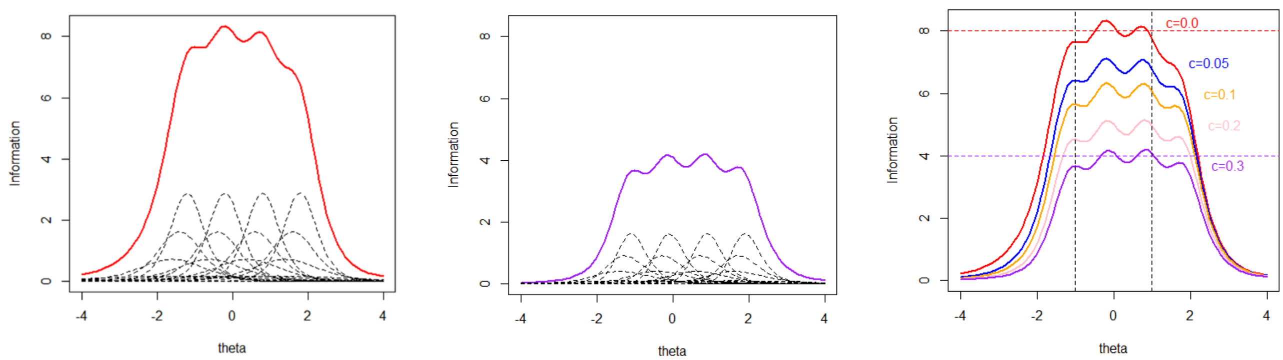
The syntax to generate the item parameters for both the left and right panels of Figure 7 is given in Listing 8 in the Appendix.

Impact on Information of Including Items with Guessing Parameter

In times when an examiner does not know about the characteristics of a list of items to be included in an assessment due to valid reasons such as time constraints, carrying out a sensitivity analysis becomes helpful in understanding the effect of including the items. In such circumstances, there

Figure 8 ■ Test & Item Information 3PL

(a) $c = 0$ **(b)** $c = 0.3$ **(c)** Varying 3PL Guessing Parameter from 0 to 0.3



is a possibility the list of items contains guessing elements that the examinees can guess correctly despite not possessing the ability to give a correct answer. The introduction of the 3PL in this circumstance is most appropriate to examine the effect of the guessing parameter on the precision of a test by carrying out a sensitivity analysis. The left panel of Figure 8 graphs the test and item information curves of an assessment by setting $c = 0$ as the base reference, varying discrimination parameters that range from 0 to 2 with an incremental of 0.5, and difficulty parameters that range from -2 to 2 with an incremental of 0.2. The middle panel of Figure 8 graphs the test information curves by setting a higher guessing parameter of 0.3 to show the difference in the value of information when compared to the left panel of Figure 8. This test information function with the higher guessing parameter of 0.3, the purple-colored curve, shows a much lower information curve. Placing these two test information curves in the right panel of Figure 8 and incorporating the guessing parameters of 0.1, 0.2, and 0.3, the distribution of the five test information functions shows that the magnitude of the information curve is greatly affected by the guessing parameters. The information difference between $c = 0$ and $c = 0.3$ is the former about twice the precision of the latter. For $c = 0.3$, the information peaks around 4 in the ability range from -1 logit to 1 logit whereas the information for $c = 0$ is around 8. In short, the effect of including a substantial number of items with high guessing parameters in an assessment could largely affect the precision, pointing out the main feature of using the 3PL model.

The syntax to generate the set of item parameters to produce the five test information functions in Figure 9 are listed in Listing 9 in the Appendix. Similar sensitivity analysis could be easily carried out by modifying the syntax with a different set of values for the difficulty and discrimination parameters and varying the guessing parameters.

Impact of Inattention Parameter using the 4PL Model

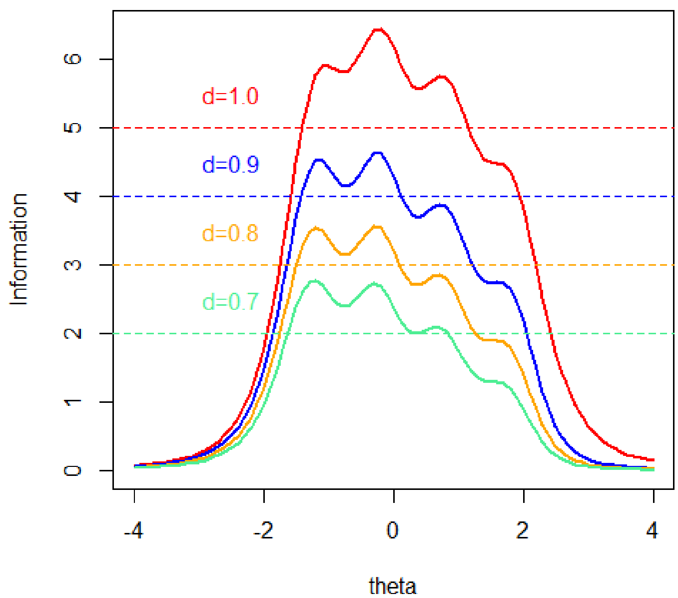
Figure 9 plots the test information function using 4PL by specifying four conditions for the inattention parameter that range from 1.0 down to 0.7 with a decremental of 0.1. It is not unexpected, given the item information function discussed earlier, the test information function of the inattention parameter that is set to 1 gives the highest information and the lowest is with the inattention parameter set to 0.7. The extent of reduction in the information is about one unit of information value with a drop of 0.1 in the inattention parameter. These variations in information in inattention parameters provide the magnitude changes in one unit of information function is about a change in 0.1 probability of the inattention parameter, however, it should be noted the conditions set for the discrimination, difficulty, and guessing parameters for this example may not apply to all conditions. An examiner may have to specify a different set of parameters according to their purposes.

The R syntax that generates the four test information functions is given in Listing 10 in the Appendix. The inattention parameter varies with a decrement of 0.1 for the four sets of 21 items from 1.0 down to 0.7 are given below.

Offsetting Effect Between Discrimination and Guessing & Inattention Parameter

As mixing all four IRT models into an assessment becomes common, the effect on information by simultaneously varying the discrimination parameter, guessing parameter, and inattention parameter becomes vital and their effects need to be more closely examined. The left panel of Figure 10 shows the magnitude of information when the discrimination parameter is set at 1.0 and with a shift of 0.1 in both the

Figure 9 ■ 4PL – Varying Inattention Parameter d from 1 to 0.7



guessing parameter and inattention parameter in the same and opposite directions. When the values of the guessing and inattention parameter are set at 0 and 1 respectively, the information is at highest. When there is an increase of the guessing parameter to 0.1 or a drop of the inattention parameter to 0.9, the information drop is about 5 information units. When there is an increase in the guessing parameter and a drop in the inattention parameter happens concurrently, the information drops further by another 5 information units.

A similar pattern in the test information could also be found in the right panel of Figure 10 when the same four sets of conditions are specified for the guessing and inattention parameters but the value of the discrimination parameter increases from 1.0 to 1.5. The information differences and gaps for both Figures are similar in values and shapes but the information for the higher discrimination specification of 1.5 is much higher than the lower discrimination of 1.0 specification. These results indicate that the information effects by varying the discrimination parameter have a larger impact in comparison to the variation of the guessing and inattention parameters. The R syntax for the item specifications is given in Listing 11 in the Appendix.

Test Length and Test Information Function

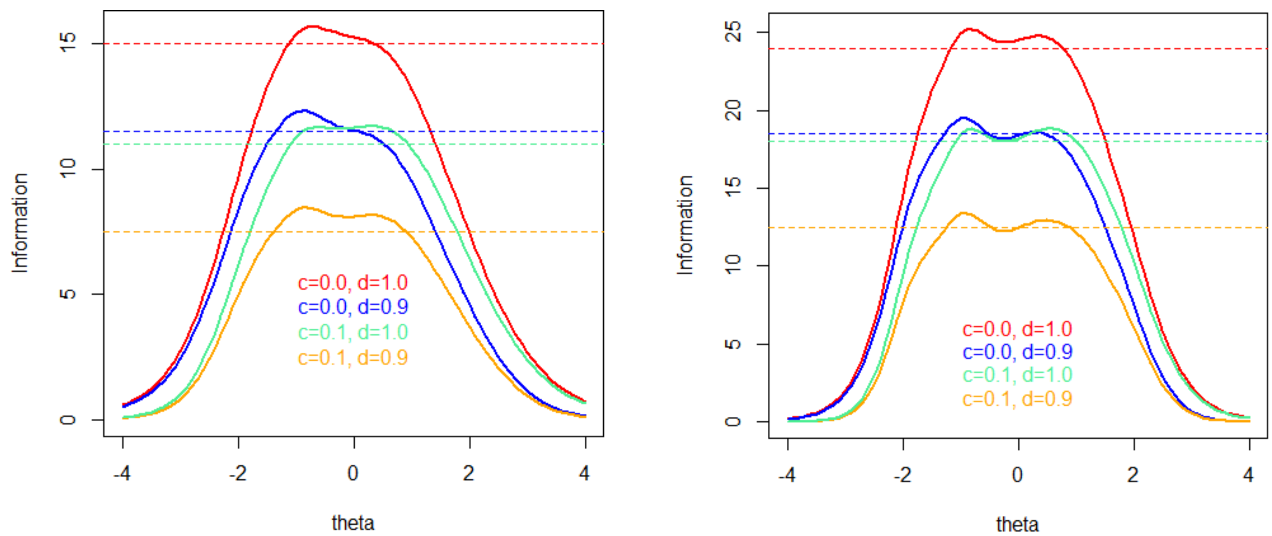
Using the test information function to set up an assessment raises a pertinent question about the number of items needed to put into an assessment that is sufficient to bring about a test with high or close to expected precision. A consensus is the longer the test, the more accurate precision will be attained (Al Kursheh et al., 2022; Brzezińska, 2020), however, the effect of the parameters and their information have to be considered to qualify this claim. Although Hambleton and Swaminathan (2013) recommend constructing the shortest possible test to meet a targeted precision via the test information function by increasing the number of test items incrementally till it reaches the targeted information, they do not examine the effect of parameters on the information that item parameters can vary the test length. Figure 11 gives a broad direction of the information effect by varying the number of items with a set of items by specifying with slight variation the discrimination, guessing, and inattention parameters.

For a set of items with discrimination, guessing, and inattention (a, c, d) parameters set as (1.5, 0, 1) and varying the number of items from 9 and increasing it to 45 items with an increment of 9 items, the top left panel of Figure 11 produces five test information curves that start with the lowest information of about 5 information units for the set of 9 items to 25 information units for the set of 45 items. These graphs show almost equal distance apart between them for the ability level between -2 and 2 logits with an increase of 5 information units by adding 9 items.

When the value of the discrimination parameter is reduced from 1.5 to 1, the top right panel of Figure 11 shows that there is a substantial reduction in the overall test infor-

Figure 10 ■ Test Information

(a) Varying c and d with $a = 1$ **(b)** Varying c and d with $a = 1.5$



mation units. Similar shapes of the test information curves and between information unit distances for the five test information curves could also be observed in the bottom left and right panels of Figure 11. The left panel of Figure 11 is when the guessing parameter is set to a higher value of 0.2 and the right left panel of Figure 11 is when the inattention parameter is set to a lower value of 0.8 respectively. These four sets of test information curves indicate that a slight change of the discrimination parameter of 0.5 has a greater effect on the information than the change in guessing and inattention parameter of 0.2 probability by varying the length of an assessment.

The syntax to generate Figure 11 is given in Listing 12 in the Appendix.

Noting for Setting An Assessment Using Information Functions

As item and test information functions are closely associated with the parameters of the four IRT models, a general recommendation and guideline that summarizes the information approach for setting up an assessment is described below.

1. Identifying the traits of the examinees is crucial for the selection of items in an assessment to meet the objective of the assessment. For instance, selecting a list of extremely difficult items has to be in line with the group of high-ability students sitting for the assessment. In short, the objective has to be aligned with the items being selected.
2. As the main concern of the 1PL model is solely on the item difficulty, the denseness of the item information over the entire ability continuum becomes the main concern in forming an assessment using 1PL.
3. The discrimination parameter is a powerful parameter that gives high item information. By aggregating a list of items with high discrimination parameters will derive a high test information function for the assessment.
4. Using items with high guessing and low inattention parameters will result in lowering the test information. However, there is a possible tradeoff between information and the inclusion of an item to attain a specific purpose. For instance, examining the level of social desirability using a high guessing parameter will result in low information but be able to capture whether participants give a social desirability answer. While one of the purposes of an assessment is not to miss out on careless examinees with high ability who answer wrongly for an item, using 4PL with an inattention parameter becomes appropriate (Liao et al., 2012; Rulison & Loken, 2009). This is especially applicable under CAT. Explaining and justifying the slipping effect is one obvious application of 4PL for high-ability students who fail to answer an easy question. Culpepper (2016, 2017) applies this model to evaluating low-stakes large-scale educational assessments. The effect of determining the level of inattention and the level of information, in practice, could turn out as a balancing issue.
5. Obtaining a cutoff point using the IRT test information

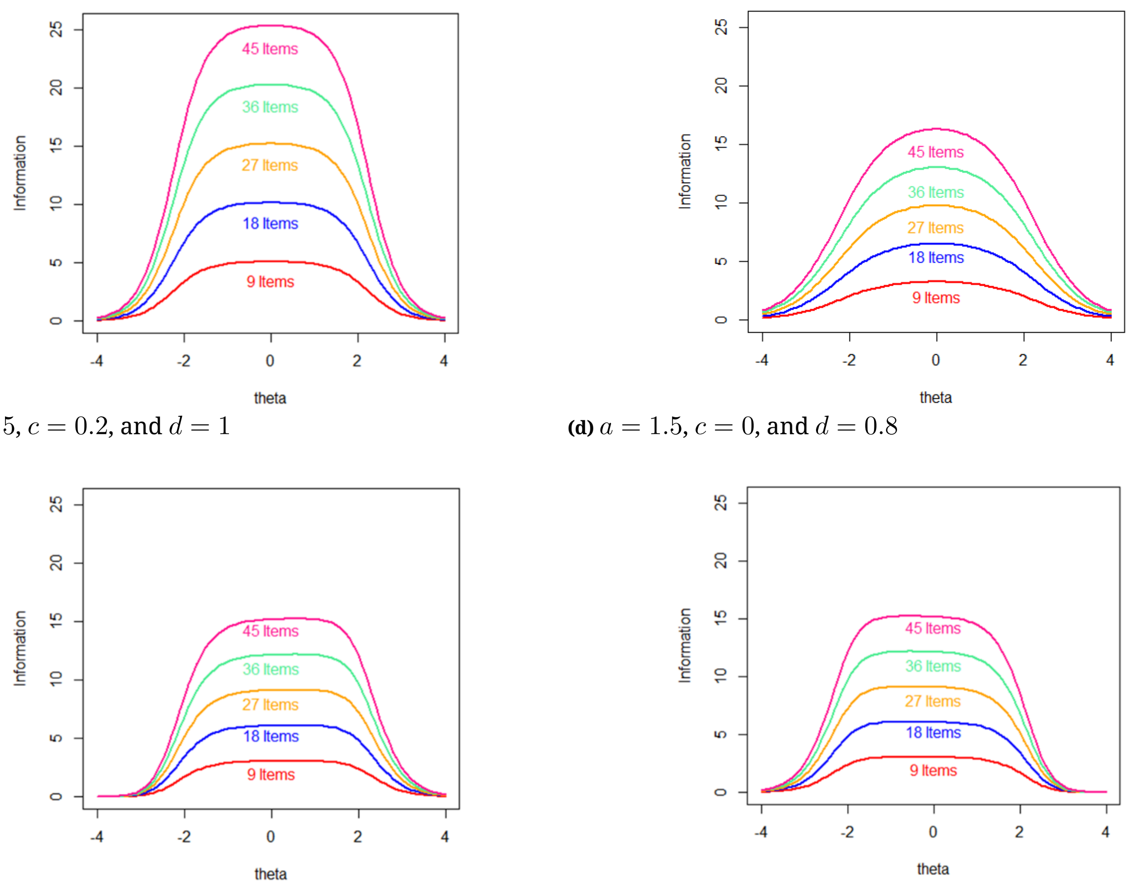
Figure 11 ■ Test Information Function

(a) $a = 1.5$, $c = 0$, and $d = 1$

(b) $a = 1$, $c = 0$, and $d = 1$

(c) $a = 1.5$, $c = 0.2$, and $d = 1$

(d) $a = 1.5$, $c = 0$, and $d = 0.8$



function is what it is capable of. For example, using the information function in constructing a test that focuses on awarding scholarships to a limited number of examines could include a list of high discrimination parameter items near the ability region. This allows for the separation of candidates with slightly higher ability than the targeted threshold to qualify for receiving the scholarship (Baker & Kim, 2017). Similarly, creating a test information function that targets specific ability levels, such as the passing score could use the same strategy.

6. The input items into an item bank probably need to include 3PL and 4PL to examine the characteristics of the items concerning their guessing and inattention.
7. Brzezińska (2020) suggests that in general, a longer test to measure an examinee's ability is more accurate than a shorter one as the test information is normally higher due to the test length. However, a lengthy test could end up fatigue with repeated contents. The IRT model and the estimated values of the items are by no means crucial factors in determining the length of the assessment to reach a targeted information level, other considerations may be significant as well.

Summary and Conclusion

Characterizing the reliability and accuracy of an assessment is one central issue in psychometric theory under IRT. In place of score fidelity and reliability, IRT offers the item and test information function respectively, showing the degree of precision at different values of the theta θ ability, (Reise & Waller, 2009). This paper first introduces the concept of item information to associate it with the item parameters. The difficulty parameter indicates the location of the maximum item information. The larger the value of the discrimination parameter, the greater the information, and the taller and narrower the information curve. The smaller the value of the guessing parameter, the greater the information, and the larger the value of the inattention parameter, the greater the information. These characteristics are helpful indicators for the selection of items in an assessment.

Several original illustrations of the information approaches are introduced in the current paper. First, it illustrates the extent of the information effects of the four binary IRT models, in particular, the effect of the 4PL inattention parameter on information. Second, it shows the offsetting information effect among the discriminating, guessing, and inattention parameters on the test information. Third, the magnitudes of these effects are examined, compared, and contrasted on their concurrent and offsetting effect. Fourth, by using the graphical approach with simple R syntax, educators and examiners can easily transport them to examine the effects of their assessments. Fifth, it sheds light that when there is a lack of information on item characteristics, sensitivity analysis can provide a way to examine the effect of information. Sixth, the association of item parameters to test length is examined, relating information to test length.

While the current emphasis of the current paper discusses the impact of information with a focus on assessment, IRT has become increasingly popular that it goes beyond education assessment. IRT has moved on to the field beyond cognitive assessment and crossed over to other disciplines in social sciences such as psychology, psychopathology, and medicine to verify psychological measures, personality assessment, clinical assessment, and other types of inventories (e.g. Cerou et al., 2020; Rouse et al., 1999; Reise & Waller, 2003, 2009; Waller & Reise, 2009; Gray-Little et al., 1997; Thomas, 2019). Reise and Waller (2003) are cautious about the terms used in assessment may not applicable to other discipline and applications. In the context of education assessment, the guessing parameter refers to the guessing to get a correct answer, however, when used in a psychological inventory, it could be referred to as an indication of social desirability (e.g. Rouse et al., 1999, reported the parameter c that ranged from .1 to .25 referring them as the level of social desirability). In this context, the magnitude of information is about the precision of the item due to the increase of social desirability.

The information approach is not the only focus on creating and forming an assessment. The selection of test items for an assessment usually goes beyond precision which could include the scope of the subject materials, content concerns, the level of the objective statement, building of distractors, the format of a test, the purpose of the evaluation, the characteristics of examinees, and the extent to which the psychometric characteristics of the test are affected by the method of correction (Al Kursheh et al., 2022; Haladyna et al., 2002; Thawabieh, 2016). Evaluation of assessment could incorporate these factors with the consideration of precision.

References

Al Kursheh, T. O., Al-Zhoon, H. S., & Al Nasraween, M. S. (2022). The effect of item form on estimating person's ability, item parameters, and information function according to item response theory (irt). *International Journal of Instruction*, *15*(3), 1111–1130. doi: 10.29333/iji.2022.15359a.

Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using r*. Springer. doi: 10.1007/978-3-319-54205-8_1.

Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *Research Bulletin*, *1981*, 81–20.

Birnbaum, A. (1958). *On the estimation of mental ability. series report no.15. project no. 7755–23*. USAF School of Aviation Medicine.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Addison-Wesley.

Brzezińska, J. (2020). Item response theory models in the measurement theory. *Communications in Statistics-Simulation and Computation*, *49*(12), 3299–3313. doi: 10.1080/03610918.2018.1546399.

Cerou, M., Peigné, S., Comets, E., & Chenel, M. (2020). Application of item response theory to model disease progression and agomelatine effect in patients with major depressive disorder. *AAPS Journal*, *22*(1), 4–4. doi: 10.1208/s12248-019-0379-x.hal-02394266.

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, *81*(4), 1142–1163. doi: 10.1007/s11336-015-9477-6.

Culpepper, S. A. (2017). The prevalence and implications of slipping on low-stakes, large-scale assessments. *Journal of Educational and Behavioral Statistics*, *42*(6), 706–725. doi: 10.3102/1076998617705653.

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

DeMars, C. (2018). Item information function. In B. B. Frey (Ed.), *The sage encyclopedia of educational research measurement, and evaluation* (pp. 899–903). Sage Publications.

Gonulates, E. (2022). *Irt: Item response theory and computerized adaptive testing functions [r package version]* (Version 0.2.7). https://CRAN.R-project.org/package=irt

Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An item response theory analysis of the rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, *23*, 443–451.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, *15*(3), 309–333. doi: 10.1207/S15324818AME1503_5.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory principles and applications*. Springer Science.

Kalkan, O. K. (2022). The comparison of estimation methods for the four-parameter logistic item response theory model. *Measurement: Interdisciplinary Research and Perspectives*, *20*(2), 93–90. doi: 10.1080/15366367.2021.1897398.

Liao, W.-W., Ho, R.-G., Yen, Y.-C., & Cheng, H.-C. (2012). The four-parameter logistic item response theory model as a robust method of estimating ability despite aberrant response. *Social Behavior and Personality*, *40*(10), 1679–1694. doi: 10.2224/sbp.2012.40.10.1679.

Moghadamzadeh, A., Salehi, K., & Khodaie, E. (2011). A comparison the information functions of the item and test on one, two and three parametric model of the item response theory (irt). *Procedia-Social and Behavioral Sciences*, *29*, 1359–1367. doi: 10.1016/j.sbspro.2011.11.374.

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

Reise, S. P., & Waller, N. G. (2003). How many irt parameters does it take to model psychopathology items? *Psychological Methods*, *8*, 164–184. doi: 10.1037/1082-989X.8.2.164.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48. doi: 10.1146/annurev.clinpsy.032408.153553.

Rouse, S. V., Finger, M. S., & Butcher, J. N. (1999). Advances in clinical personality measurement: An item response theory analysis of the mmpi–2 psy–5 scales. *Journal of Personality Assessment*, *72*, 282–307. doi: 10.1207/S15327752JP720212.

Rulison, K., & Loken, E. (2009). I've fallen and i can't get up: Can high-ability students recover from early mistakes in cat? *Applied Psychological Measurement*, *33*, 83–101. doi: 10.1177/0146621608324023.

SAS/STAT. (2023). *Bayesian irt models: Unidimensional bianry models*. https://support.sas.com/rnd/app/stat/examples/irt/irt.htm

Thawabieh, A. M. (2016). A comparison between two test item formats: Multiple-choice items and completion items. *British Journal of Education*, *4*(8), 63–74.

Thomas, M. L. (2019). Advances in applications of item response theory to clinical assessment. *Psychological Assessment*, *31*(12), 1442–1455. doi: 10.1037/pas0000597.

Waller, N. G., & Reise, S. P. (2009). *Measuring psychopathology with non-standard irt models: Fitting the four-parameter model to the mmpi embretson, s., roberts, j.* S. New directions in psychological measurement with model-based approaches.Washington.

Yao, L. (2019). Item selection methods for computer adaptive testing with passages. *Frontiers in Psychology*, *10*(240), 1–19. doi: 10.3389/fpsyg.2019.00240.

Zheng, C., Guo, S., & Kern, J. L. (2021). Fast Bayesian estimation for the four-parameter logistic model (4plm). *SAGE Open*, *12*, 1–13. doi: 10.1177/21582440211052556.

Appendix: Listings

Listing 1:

```
library(irt)
Item01 <- itempool(b = seq(-2,2,1), D = 1.7)
irt::plot_info(Item01,
      title="Item Information Function - 1PL\nVarying Parameter b")
```

Listing 2:

```
Item02 <- itempool(b = c(rep(-1,3),rep(1,3)),
                   a = c(seq(1.5,3.5,1),seq(1.5,3.5,1)),
                   D = 1.7)
irt::plot_info(Item02, theta_range = c(-2, 2),
      title="Item Information Function\n2PL, Varying Parameter a")
```

Listing 3:

```
thetaRange <- seq(from=-2, to=2,by=1)
ItemInfo02 <- info(ip=Item02,theta=thetaRange)
rownames(ItemInfo02) <- seq(from=-2, to=2,by=1)
round(ItemInfo02,4)
##    Item_1 Item_2 Item_3 Item_4 Item_5 Item_6
## -2 0.4368 0.2505 0.0918 0.0031 0.0001 0.0000
## -1 1.6256 4.5156 8.8506 0.0392 0.0037 0.0002
## 0  0.4368 0.2505 0.0918 0.4368 0.2505 0.0918
## 1  0.0392 0.0037 0.0002 1.6256 4.5156 8.8506
## 2  0.0031 0.0001 0.0000 0.4368 0.2505 0.0918
```

Listing 4:

```
Item03 <- itempool(b = rep(0,7),
                   a = rep(1,7),
                   c = seq(0.0, 0.30, 0.05),
                   D = 1.7)
irt::plot_info(Item03, theta_range = c(-3, 3),
      title = "Item Information Function\n3PL, Varying Parameter c")
```

Listing 5:

```
Item04 <- itempool(b = rep(0,7),
                   a = rep(1,7),
                   c = rep(0,7),
                   d = seq(0.7, 1.00, 0.05),
                   D = 1.7)
irt::plot_info(Item04, theta_range = c(-3, 3),
      title="Item Information Function\n4PL, Varying Parameter d")
```

Listing 6:

```
plot(theta,Test1_Info,
     type="l",lty=1,ylab="Information",
     main="Test and Item Information Functions - 1PL\nb vary from -2 to 2 with
   incremental 0.5",
     ylim=c(0,max(Test1_Info)),col="red",lwd=2)
for (i in 1:(nrow(Item1_Info)))
  lines(theta,Item1_Info[, i],type="l",lty=2,lwd=1.5)
for (i in 2:(nrow(Item1_Info)-1)){
  lines(theta,rowSums(Item1_Info[, 1:i]),type="l",lty=2)
}
```

Listing 7:

```
Item2 <- itempool(b = c(seq(-2,-1,0.2),seq(1,2,0.2)), D = 1.7)
Item2_Info <- info(Item2,theta)
Test2_Info <- rowSums(Item2_Info)
plot(theta,Test2_Info,
     type="l",lty=1,ylab="Information",
     main="1PL - Two Distinct Ability Groups\nTest and Item Information Functions",
     ylim=c(0,max(Test2_Info)),col="red",lwd=2)
for (i in 1:(nrow(Item2_Info)))
  lines(theta,Item2_Info[, i],type="l",lty=2,lwd=1.5)
```

Listing 8:

```
# Figure 7 Left Panel
Item3 <- itempool(b = c(seq(-2,-1,0.2),seq(1,2,0.2)),
                  a = c(seq(0,2.2,0.2)),
                  D = 1.7)
# Figure 7 Right Panel
Item4 <- itempool(b = c(seq(2,1,-0.2),seq(-1,-2,-0.2)),
                  a = c(seq(0,2.2,0.2)),
                  D = 1.7)
```

Listing 9:

```
itempool(b = seq(-2,2,0.2),
    a = c(rep(seq(0,2,0.5),4),0),
    c = rep(0,21),
    D = 1.7)
itempool(b = seq(-2,2,0.2),
    a = c(rep(seq(0,2,0.5),4),0),
    c = rep(0.05,21),
    D = 1.7)
itempool(b = seq(-2,2,0.2),
    a = c(rep(seq(0,2,0.5),4),0),
    c = rep(0.1,21),
    D = 1.7)
itempool(b = seq(-2,2,0.2),
    a = c(rep(seq(0,2,0.5),4),0),
    c = rep(0.2,21),
    D = 1.7)
itempool(b = seq(-2,2,0.2),
    a = c(rep(seq(0,2,0.5),4),0),
    c = rep(0.3,21),
    D = 1.7)
```

Listing 10:

```
itempool(b = seq(-2,2,0.2),
         a = c(rep(seq(0,2,0.5),4),0),
         c = c(rep(seq(0,0.3,0.1),5),0),
         d = rep(1,21)
         D = 1.7)
itempool(b = seq(-2,2,0.2),
         a = c(rep(seq(0,2,0.5),4),0),
         c = c(rep(seq(0,0.3,0.1),5),0),
         d = rep(0.9,21),
         D = 1.7)
itempool(b = seq(-2,2,0.2),
         a = c(rep(seq(0,2,0.5),4),0),
         c = c(rep(seq(0,0.3,0.1),5),0),
         d = rep(0.8,21),
         D = 1.7)
itempool(b = seq(-2,2,0.2),
         a = c(rep(seq(0,2,0.5),4),0),
         c = c(rep(seq(0,0.3,0.1),5),0)
         d = rep(0.7,21),
         D = 1.7)
```

Listing 11:

```
itempool(b = seq(-2,2,0.1),
         a = rlnorm(41, 0, .3),
         c = rep(0,41),
         d = rep(1,41),
         D = 1.7)
itempool(b = seq(-2,2,0.1),
         a = rlnorm(41, 0, .3),
         c = rep(0.0,41),
         d = rep(0.9,41),
         D = 1.7)
itempool(b = seq(-2,2,0.1),
         a = rlnorm(41, 0, .3),
         c = rep(0.1,41),
         d = rep(0.9,41),
         D = 1.7)
itempool(b = seq(-2,2,0.1),
         a = rlnorm(41, 0, .3),
         c = rep(0.1,41),
         d = rep(1.0,41),
         D = 1.7)
```

Listing 12:

```
itempool(b = seq(-2,2,0.5),
         a = rep(1.5,9),
         c = rep(0.0,9),
         d = rep(0.8,9),
         D = 1.7)
itempool(b = rep(seq(-2,2,0.5),2),
         a = rep(1.5,18),
         c = rep(0,18),
         d = rep(0.8,18),
         D = 1.7)
itempool(b = rep(seq(-2,2,0.5),3),
         a = rep(1.5,27),
         c = rep(0.0,27),
         d = rep(0.8,27),
         D = 1.7)
itempool(b = rep(seq(-2,2,0.5),4),
         a = rep(1.5,36),
         c = rep(0.0,36),
         d = rep(0.8,36),
         D = 1.7)
itempool(b = rep(seq(-2,2,0.5),5),
         a = rep(1.5,45),
         c = rep(0.0,45),
         d = rep(0.8,45),
         D = 1.7)
```

Citation

Tan, T. K. (2024). Evaluating assessment via item response theory utilizing information function with R. *The Quantitative Methods for Psychology*, *20*(1), 33–49. doi: 10.20982/tqmp.20.1.p033.

Received: 07/11/2023 ∼ Accepted: 17/02/2024