

Whelan, T. J., Stoughton, J. W., Craig, S. B., & Parry, K. W. (2014, May). *A short-form of the Perceived Leadership Integrity Scale (sPLIS)*. Paper to be presented at the 29th annual conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

## A short-form of the Perceived Leadership Integrity Scale (sPLIS)

Thomas J. Whelan, J. William Stoughton, and S. Bartholomew Craig  
North Carolina State University

Parry, K. W.  
Bond University

A sample of employees completed the full Perceived Leadership Integrity Scale (PLIS) designed to measure subordinate perceptions of their direct boss' integrity. Statistical analysis was used to isolate a unidimensional factor structure demonstrating that all questions surveyed the same theoretical domain. This factor structure was confirmed with CFA, another powerful statistical technique. A short-form version of the PLIS was created removing items based on guidelines in the literature. Following this, IRT parameters were estimated from a second field sample to assess how sensitive the short-form was to subordinate perceptions. This should help researchers create shorter surveys that keep participants' attention.

The use of self-report survey instruments has been critical to the development of the psychological and organizational sciences. Self-report survey measures are used in: selection and development contexts (Morgeson et al., 2007; Rothstein & Goffin, 2006), employee attitude and opinion assessments (Thompson & Surface, 2007), and traditional higher-education psychological measurement (Rogelberg et al., 2003), to name just a few. These organizational survey efforts rely upon the effortful attention, engagement, and affective reactions of participants (Meade & Craig, 2012; Schwarz, 1999; Thompson, Surface, Martin, & Sanders, 2003; Tourangeau, Rips, & Rasinski, 2000) to ensure success and completion in both cross-sectional (i.e., attrition) and longitudinal (i.e., nonresponse) survey efforts (Donnellan, Oswald, Baird, & Lucas, 2006).

Previous research has demonstrated that the length of a survey instrument correlates with survey attrition (Meade & Craig, 2012). Even when working with a highly motivated sample (e.g., job applicants) participants are likely to feel fatigue effects from longer surveys (Berry et al., 1992). Survey attrition and/or survey nonresponse in longitudinal survey efforts threatens the statistical power of organizational and psychological measurement efforts as sample size decreases (Rogelberg et al., 2003; Shadish, Cook, & Campbell, 2002). However, attrition and nonresponse are not the only risks presented by longer surveys. The fatigue likely experienced by participants has the potential to create serious psychometric issues as participants respond carelessly in longer surveys (Meade & Craig, 2012). This careless or inattentive responding constitutes error variance, which may serve to attenuate correlations, reduce

estimates of reliability, and potentially cause specious factor analytic results (Meade & Craig, 2012). When the survey is being administered for scale development purposes these issues may be especially problematic as scale development efforts rely on item intercorrelations (Hinkin, 1998; Meade & Craig, 2012).

Classical Test Theory (CTT), the basis of most organizational and psychological researchers “rules” for measurement, relies on longer instruments for scale reliability (Embretson & Reise, 2000; Nunnally & Bernstein, 1994). This operates in opposition to the outcomes presented above, where longer measurement instruments create an environment for undesirable measurement consequences. The need for short-forms of longer scales exists. Indeed, when psychometric properties are not sacrificed in a short-form scale the efficiency of the shorter instrument is attractive (Gosling, Rentfrow, & Swann, 2003). Accordingly, the purpose of this manuscript is twofold, (1) develop a short-form version of the Perceived Leader Integrity Scale (PLIS) and (2) elucidate the process of short-form scale refinement using Item Response Theory (IRT).

### **The Measurement of Leader Integrity**

Leader integrity has been linked to a number of positive outcomes. For example, Vitell and Davis (1990) found a positive correlation between employees’ perceptions of leader integrity and employee job satisfaction. In fact, Posner and Schmidt (1984) have gone as far to say that some managers respect integrity above all other values, even that of competence in other prospective managers.

Psychometrically sound measurement of leader integrity was initiated with the Perceived Leader Integrity Scale (PLIS) developed and validated by

Craig and Gustafson (1998). Previous attempts at measuring leader integrity (e.g., Vitell & Davis, 1990) inadequately sampled the content domain employing instruments of one to three items. Single item measures are often inadequate for sampling the content domain, but also do not provide information for managers and leaders as tools for development. The majority of items utilized in the PLIS focus on the interpersonal relationship between leaders and subordinates. This was done in order to keep the scale relevant in a wide-variety of organizational settings and to allow the scale to be used as a development tool for leaders (Craig & Gustafson, 1998).

Items were initially generated for the PLIS from a rule-based utilitarian approach, which labels acts as immoral or unethical if the consequences of the act violate explicit rules or norms that maximize the good of all; seven behavioral domains that were consistent with these guidelines: (a) training and development, (b) resource/workload allocation, (c) truth-telling, (d) unlawful discrimination, (e) compliance with policies and procedures, (f) maliciousness, and (g) self-protection (Craig & Gustafson, 1998). The final PLIS scale contained a total of 31 items. One goal of this manuscript is to derive a short-form Perceived Leader Integrity Scale (or sPLIS) from the original long-form 31 item scale; however, we have two goals for our analysis, (a) to permit assessment of attributions of integrity consistent with the PLIS and (2) to retain enough discrete behavioral information that can be used to give feedback to the managers or supervisors who are targets of evaluation as with the original long-form, which was a marked improvement over the previous one item measures of leader integrity.

### **Short-form Scale Development**

Short-form individual differences scales are not without drawbacks. For instance, short-forms may lack power to discriminate between individuals on latent personality constructs (Woods & Hampson, 2005). Short-forms created under the traditional scale development practices of classical test theory also may fail to generalize across independent samples, and may not have well understood psychometric properties. As such, short-form scales can be seen as limited in application and prone to potential problems associated with accurate measurement of the latent personality constructs they are purported to assess.

An alternative approach to creating short-forms of individual differences scales is the use of item response theory (IRT). IRT is a model-based theory of measurement that considers how item parameters and an individual's standing on a latent construct ( $\theta$ ; referred to as *theta*) relate to an individual's responses to the items representing that construct (Embretson & Reise, 2000). The advantage of IRT in the context of short-form individual differences scale development is that it permits item characteristics to be assessed independently from the subjects who have responded to the items, thereby permitting estimation of item parameters that are not sample dependent. Said another way, the relationship of items to the underlying latent construct they measure can be gauged without relying on the relative standing of subjects on that latent construct.

There is some precedent for using IRT-based methods to examine the psychometric properties of scales that measure individual differences (e.g., Childs, Dahlstrom, Kemp, & Panter, 2000; Ferrando, 2001; Ferrando & Anguiano-Carrasco, 2010). For instance, Rauch, Schweizer, and Moosbrugger (2008) employed Samejima's (1969) graded response model to demonstrate the

effectiveness of an optimism scale. Scherbaum, Cohen-Charash, and Kern (2006) compared three measures of general self-efficacy on the basis of the relative strength of the item parameters in each measure. Further, IRT has been used in recent research to create several short-form measures from what were originally longer scales. Thorpe, McMillan, Sigmon, Owings, Dawson, and Bouman (2007) described the use of IRT to generate a four-item screening tool for irrationality. Tracey (2010) utilized IRT to reduce a long form vocational interest inventory totaling over 340 items to a short-form of 80 items and demonstrated that the short-form retained most psychometric properties of the long form instrument.

## Method

### Participants and Design

The first sample comes from Craig and Gustafson (1998). Survey packets were distributed to all 826 full-time and part-time employees of the College of Arts and Sciences at a large southeastern university. 296 usable surveys were returned, for a response rate of 36.2%. Of respondents in this sample, 57% were male and 43% were female.

The second sample comes from Parry and Proctor-Thomson (2002). 28 of the original PLIS items were distributed to 6,025 managers throughout New Zealand from a range of both public and private sector organizations. 1,351 usable surveys were returned for a response rate of 22.5%. Of respondents in this sample, 77.6% were male and 22.4% were female.

### Measures

Between 77 and 28 items (see Craig & Gustafson, 1998, and Parry & Proctor-

Thomson, 2002, respectively) were administered to measure perceived leader integrity. For each item, participants were asked to rate the extent to which the statement described their immediate supervisor. Response options were presented on a 4-point Likert-style scale ranging between 1 (“Not at all”) and 5 (“Exactly”), and reverse-scored such that higher values indicated greater reported levels of perceived leader integrity.

## Results

EFA models were tested and reliability analysis conducted using SPSS. Subsequent CFA models were tested using Mplus version 5 software (Muthén & Muthén, 2007).

**Exploratory Factor Analysis (EFA).** We conducted an EFA on the 27 of the PLIS items common to both samples, utilizing the first sample indicated above from Craig and Gustafson (1998). A quartimin rotation suggested the presence of four factors per the Kaiser criterion with eigenvalues of 16.63, 1.79, 1.14 and 1.04, implying that the items did not form a single construct. Using the criterion for retaining an item based on a factor loading greater than 0.40 suggested by Hinkin (1998), four items were deleted for loading on other factors. A subsequent EFA of the remaining items yielded a single factor with an eigenvalue of 69.59, accounting for 68.11% of the variance.

**Reliability Analysis.** The 23 items remaining yielded a reliability coefficient of  $\alpha = 0.98$ . An analysis of the impact on reliability if items were deleted resulted in the deletion of two more items with no impact to the coefficient value.

**Confirmatory Factor Analysis (CFA).** Next, the 21 PLIS items retained from the EFA were subjected to a CFA with all items loading onto a single latent

variable. The model yielded poor fit,  $\chi^2(324) = 2011.85$ ,  $p < 0.001$ , CFI = 0.80, TLI = 0.78, RMSEA = 0.13, SRMR = 0.06 (see Hu & Bentler, 1999, for a discussion of cutoff criteria for approximate fit indices). Based on these unsatisfactory results, the items were subjected to further trimming based on modification indices, as suggested by Hinkin (1998). An additional 12 items were removed and the model respecified. The model yielded adequate fit,  $\chi^2(27) = 59.62$ ,  $p > 0.001$ , CFI = 0.99, TLI = 0.98, RMSEA = 0.06, SRMR = 0.02.

Descriptive statistics and intercorrelations for the final list of items in the sPLIS can be found in Table 1. After establishing unidimensionality by demonstrating a measurement model of acceptable fit, a reliability coefficient was calculated for the final set of nine items. The resulting value ( $\alpha = 0.95$ ) indicated that the scale displays adequate internal consistency.

**Information Parameters.** Similar to Craig and Gustafson (1998), we utilized MULTILOG (Thissen, 1991) to estimate item information parameters. Samejima’s (1969) Graded Response model, which is an item response theory (IRT) model that is commonly used to estimate item parameters with polytomous items such as the forced-choice sPLIS items.

The concept of test information in IRT is important due to its relationship to the standard error of measurement, such that a test is more psychometrically precise at ranges of the latent trait where test information is highest. Item information, which can be seen as additive within a set of unidimensional items, can be seen as a transformation of IRT item parameters such as discrimination (Embretson & Reise, 2000).

The test information function for the sPLIS is presented in Figure 1; item information functions for the nine sPLIS items are presented in Figures 2-10. As can

be seen, the sPLIS test information lies primarily between scale scores of 0 and 3, with the preponderance of information between 1 and 2. Being reverse-scored from the wording directionality of the item stems, this suggests that the sPLIS is better suited to distinguishing leaders displaying fewer negative characteristics. Said another way, the sPLIS is more sensitive to measuring leaders who exhibit more integrity behaviors to subordinates, and has a higher standard error of measurement in distinguishing leaders displaying low integrity.

### **Discussion**

The need for short-forms of longer scales exists, both from a research standpoint of parsimony of measurement and latent trait definitions as well as a practitioner need for economical data collection in modern organizations. Clearly, when psychometric properties are not sacrificed in a short-form scale, the efficiency of the shorter instrument can be attractive (Gosling et al., 2003). The purpose of this paper was to create a short-form of the PLIS, a scale designed to capture global attributions of leader integrity by a subordinate employee. In defining the sPLIS, the resulting scale displayed a unidimensional factor structure, adequate internal reliability, in addition to IRT-based information functions to inform potential consumers of the sPLIS as to the scale's likely sensitivity when used in an applied setting.

### **Implications**

The implications for use of the sPLIS cross both research and practice. For research in organizations, this short-form of the PLIS can be used when researching related topics such as justice perceptions or job satisfaction with less risk of respondent

fatigue. The sPLIS should make it easier to include the construct of leader integrity attributions as a part of research with executive and managerial samples, which are typically challenging to survey (Cycyota & Harrison, 2006).

In practice, the sPLIS can be used as part of a battery of scales in a 360-degree feedback evaluation or other developmental efforts. The relatively short length of the scale, while losing some specificity in the range of behavioral feedback compared to Craig and Gustafson's (1998) full PLIS measure, still allows a range of concepts to be discussed with feedback targets. For instance, endorsement of global attribution items such as "Is evil" or "Always gets even" could prompt a different discussion with a feedback target than endorsement of a behavioral item such as "Deliberately fuels conflict among employees."

### **Future Research**

The original PLIS was written with three types of foci across all items: global attribution items (e.g. "is evil"), behavior-toward-rater items ("lies to me"), and behavior-toward-others items ("lies to other people"). In the course of the long-form scale development, the behavior-toward-others items were deleted because they didn't contribute uniquely beyond the behavior-toward-rater items. Future research may seek to make short-form scales out of the larger pool of original items in order to permit measurement of leader integrity perceptions targeted more specifically to the domain of behavior-toward-rater items. Further, combinations of items that yield information functions at lower scale scores could be identified to align to diagnostic versus developmental purposes, such as identifying problem supervisors as opposed to identifying potential leaders with higher integrity attributions.

## References

- Berry, D. T. R., Wetter, M. W., Baer, R. A., Larsen, L., Clark, C., & Monroe, K. (1992). MMPI-2 random responding indices: Validation using a self-report methodology. *Psychological Assessment, 4*(3), 340–345. doi:10.1037/1040-3590.4.3.340
- Craig, S. B., & Gustafson, S. B. (1998). Perceived leader integrity scale: An instrument for assessing employee perceptions of leader integrity. *The Leadership Quarterly, 9*(2), 127–145. doi:10.1016/S1048-9843(98)90001-7
- Cycyota, C. S., & Harrison, D. A. (2006). What (not) to expect when surveying executives: A meta-analysis of top manager response rates and techniques over time. *Organizational Research Methods, 9*, 133-160.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192–203. doi:10.1037/1040-3590.18.2.192
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1
- Hinkin, T. R. (1998). A Brief Tutorial on the Development of Measures for Use in Survey Questionnaires. *Organizational Research Methods, 1*(1), 104–121. doi:10.1177/109442819800100106
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological methods, 17*(3), 437–55. doi:10.1037/a0028085
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683–729. doi:10.1111/j.1744-6570.2007.00089.x
- Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York, NY: McGraw-Hill.
- Parry, K., & Proctor-Thomson, S. B. (2002). Perceived integrity of transformational leaders in organizational settings. *Journal of Business Ethics, 35*, 75-96.
- Posner, B. Z., & Schmidt, W. H. (1984). Values and the American manager: An update. *California Management Review, 26*, 202–216.
- Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmuller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology, 88*(6), 1104–14. doi:10.1037/0021-9010.88.6.1104

- Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*(2), 155–180. doi:10.1016/j.hrmr.2006.03.004
- Samejima, F. (1969). Estimation of latent ability using a response pattern of latent scores. *Psychometrika Monograph*, no. 17.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*(2), 93–105. doi:10.1037/0003-066X.54.2.93
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.
- Thissen, D. (1991). *MULTILOG user's guide*. Mooresville, IN: Scientific Software.
- Thompson, L. F., & Surface, E. A. (2007). Employee surveys administered online: Attitudes toward the medium, nonresponse, and data representativeness. *Organizational Research Methods*, *10*(2), 241–261. doi:10.1177/1094428106294696
- Thompson, L. F., Surface, E. A., Martin, D. L., & Sanders, M. G. (2003). From paper to pixels: Moving personnel surveys to the web. *Personnel Psychology*, *56*(1), 197–227. doi:10.1111/j.1744-6570.2003.tb00149.x
- Tourangeau, R. L., Rips, L. J., & Rasinski, K. A. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.
- Vitell, S. J., & Davis, D. L. (1990). The relationship between ethics and job satisfaction: An empirical investigation. *Journal of Business Ethics*, *9*(6), 489–494. doi:10.1007/BF00382842

Table 1

*Means, Standard Deviations, and Intercorrelations for sPLIS Items*

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. Always gets even.	1.47	0.87	--							
2. Lies to me.	1.45	0.84	0.69	--						
3. Would risk me to protect himself/herself in work matters.	1.66	1.00	0.74	0.72	--					
4. Deliberately fuels conflict among employees.	1.33	0.72	0.72	0.70	0.70	--				
5. Is evil.	1.19	0.58	0.64	0.69	0.60	0.62	--			
6. Has it in for me.	1.23	0.67	0.70	0.65	0.66	0.62	0.65	--		
7. Is a hypocrite.	1.47	0.85	0.76	0.78	0.75	0.68	0.69	0.70	--	
8. Would risk me to get back at someone else.	1.21	0.54	0.73	0.70	0.70	0.69	0.64	0.64	0.74	--
9. Would fire people just because (s)he doesn't like them if (s)he could get away with it.	1.40	0.84	0.71	0.63	0.69	0.59	0.61	0.70	0.74	0.70

*Note:* all correlations are significant at  $p < .0001$ ,  $N = 296$ .

Table 2

*Factor Loadings and Standard Errors for sPLIS Items*

Item Stem	$\lambda$	SE
1. Always gets even.	0.858	0.017
2. Lies to me.	0.840	0.019
3. Would risk me to protect himself/herself in work matters.	0.838	0.019
4. Deliberately fuels conflict among employees.	0.799	0.023
5. Is evil.	0.769	0.026
6. Has it in for me.	0.798	0.023
7. Is a hypocrite.	0.898	0.013
8. Would risk me to get back at someone else.	0.855	0.018
9. Would fire people just because (s)he doesn't like them if (s)he could get away with it.	0.813	0.021

*Note:*  $N = 296$ .

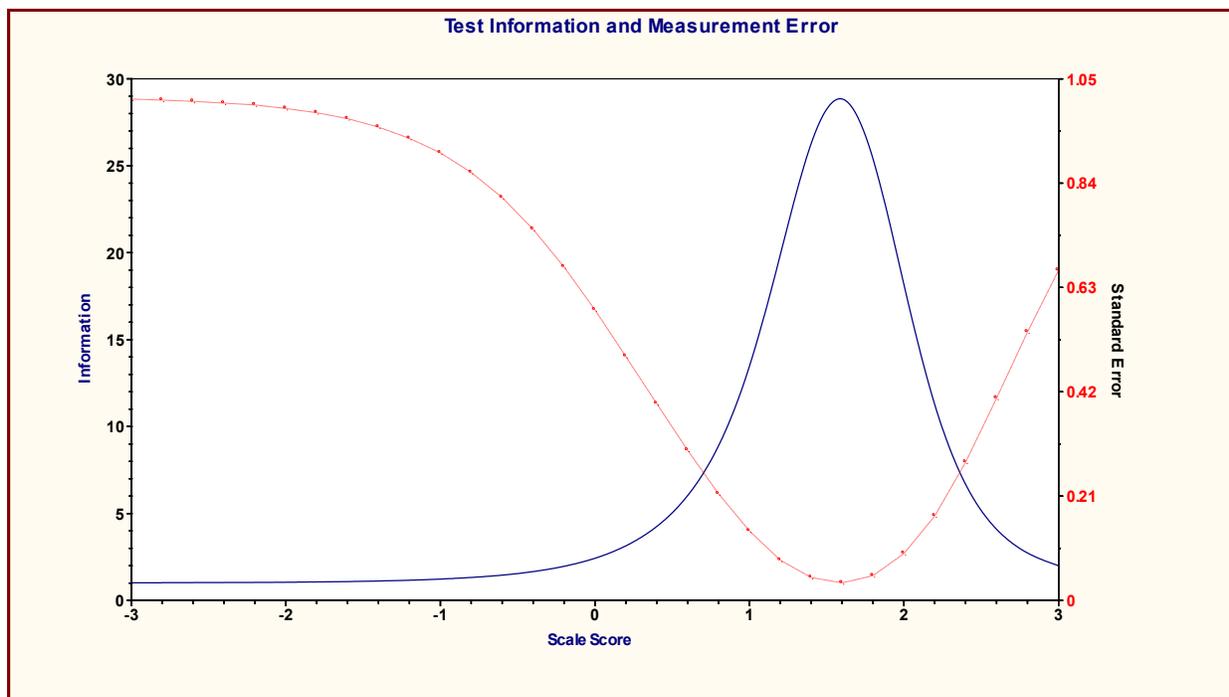
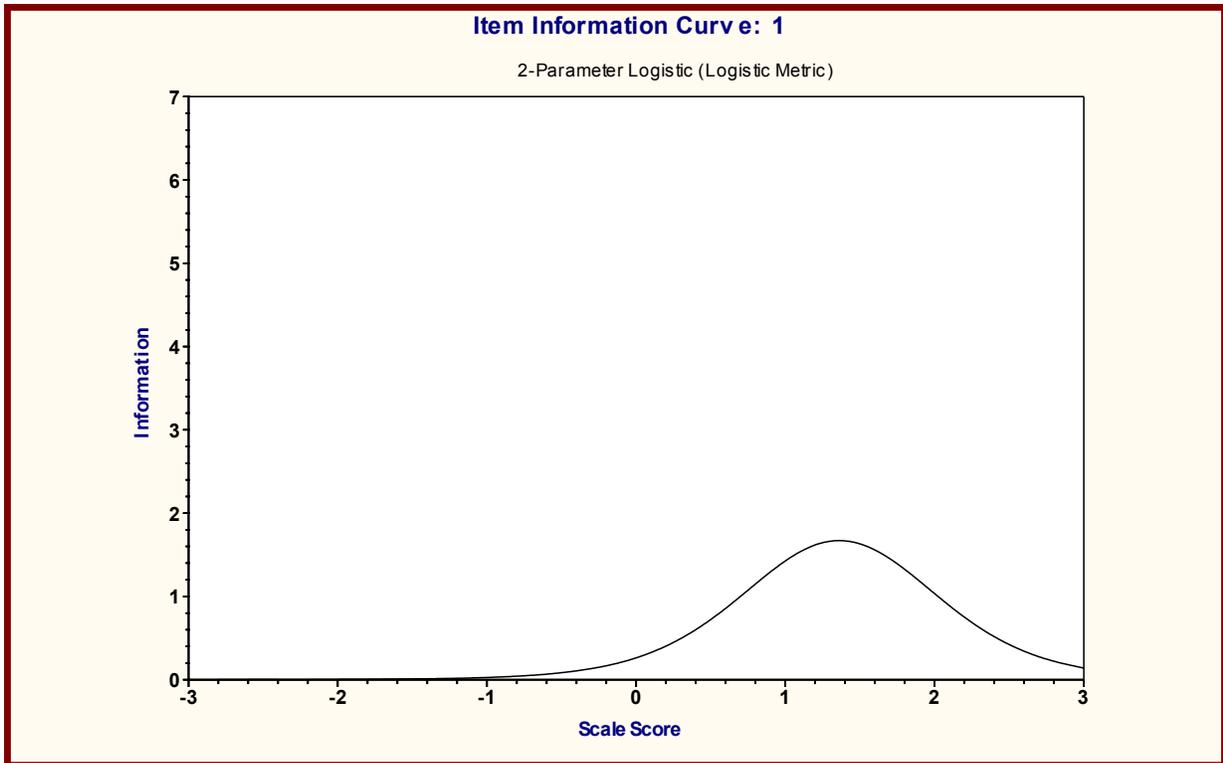
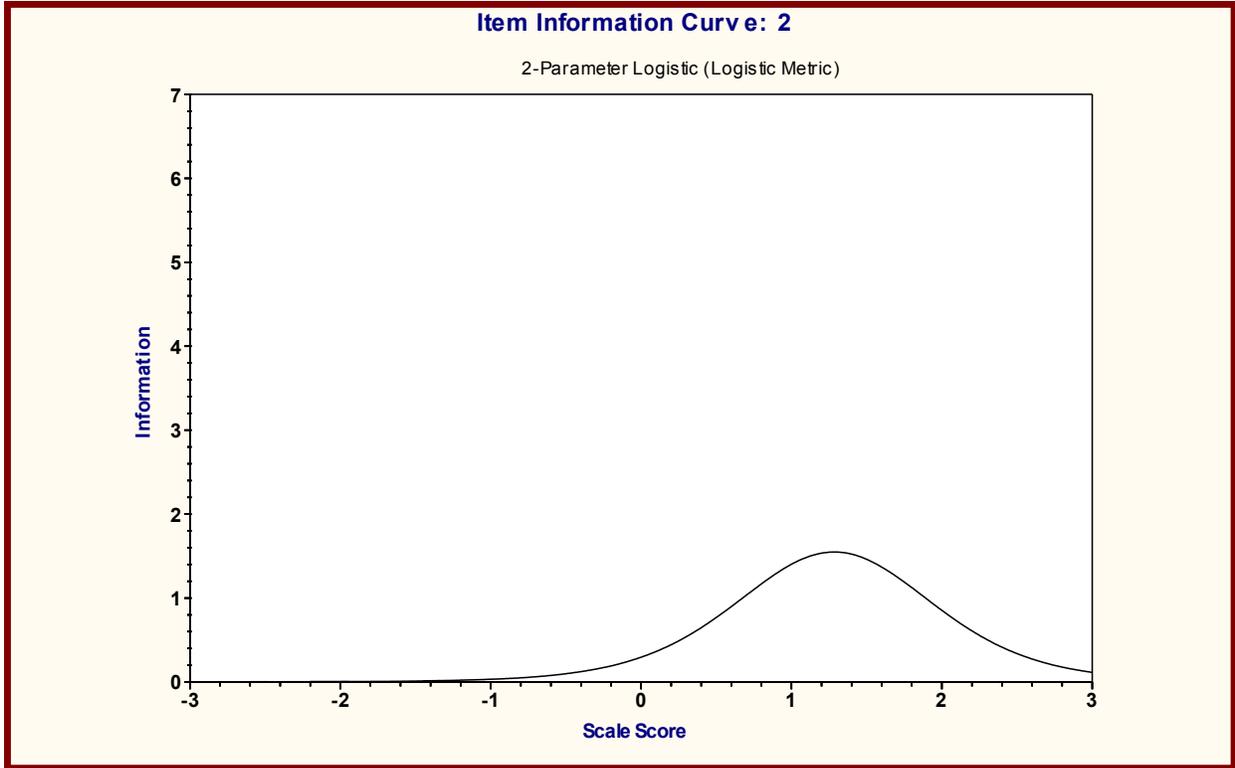


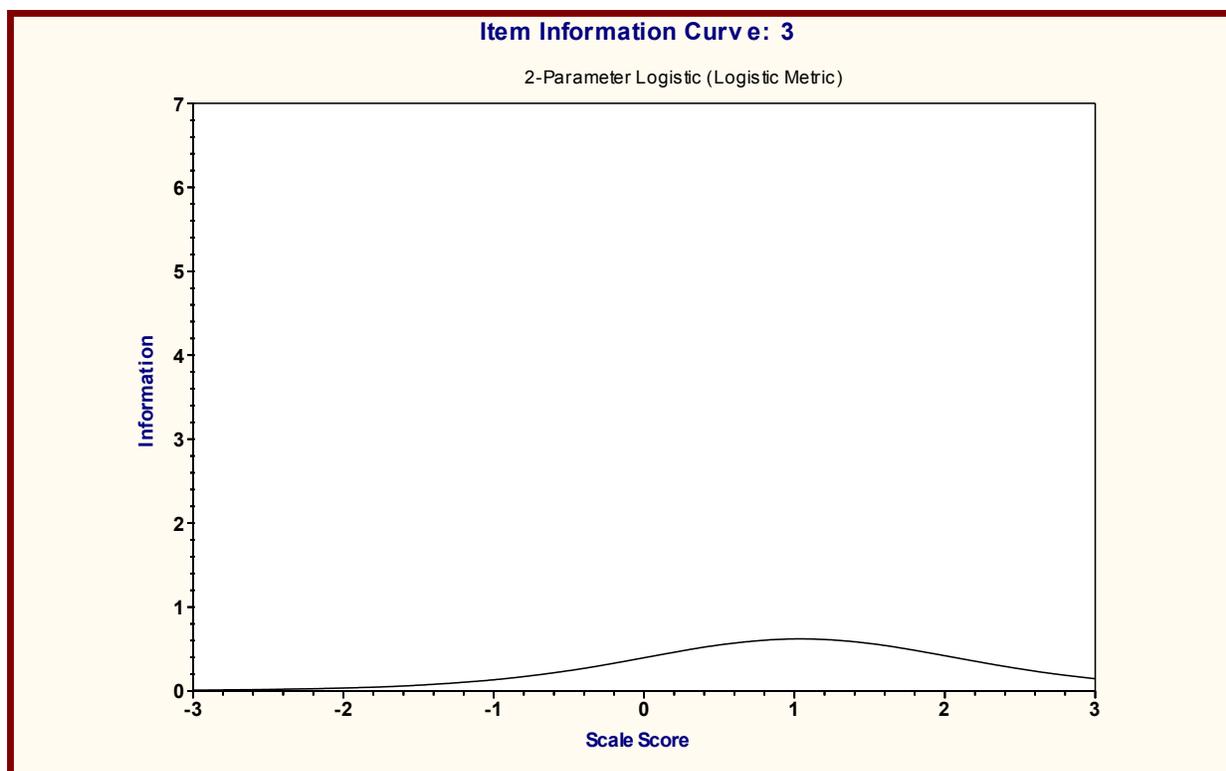
Figure 1. Test Information Function for sPLIS



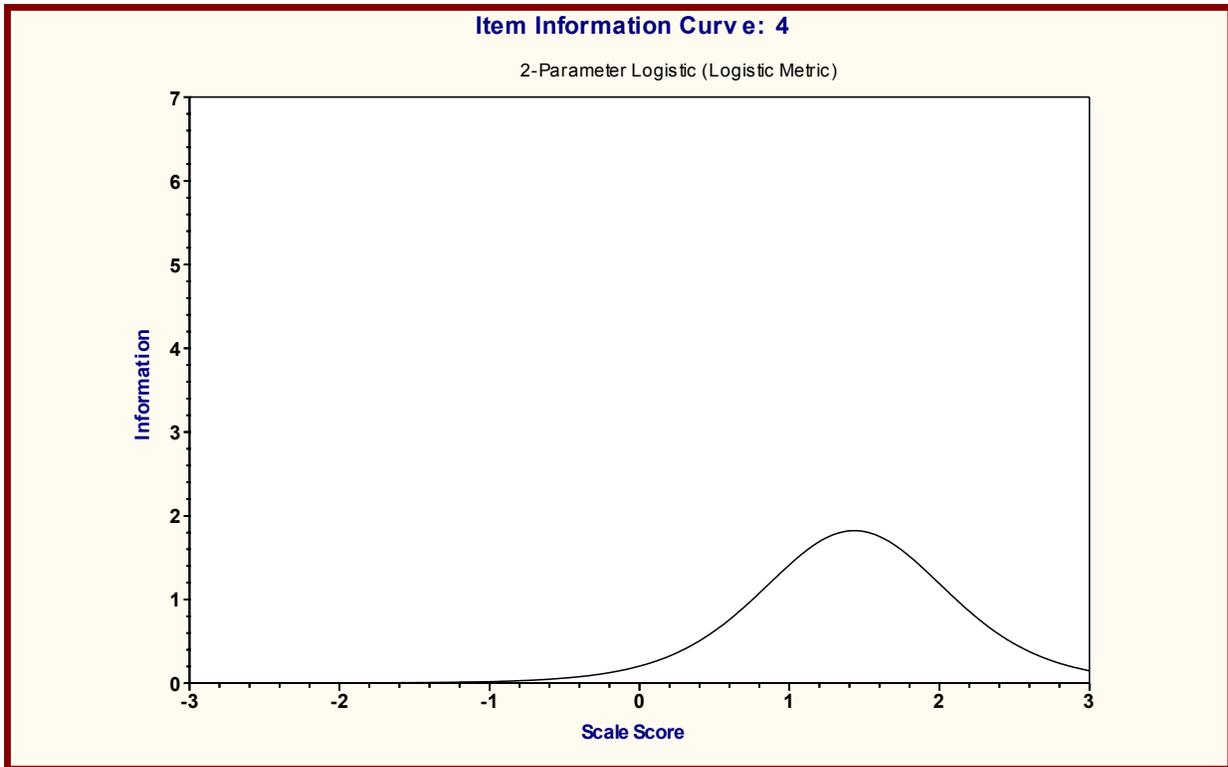
*Figure 2.* Item Information Function for “Always gets even”



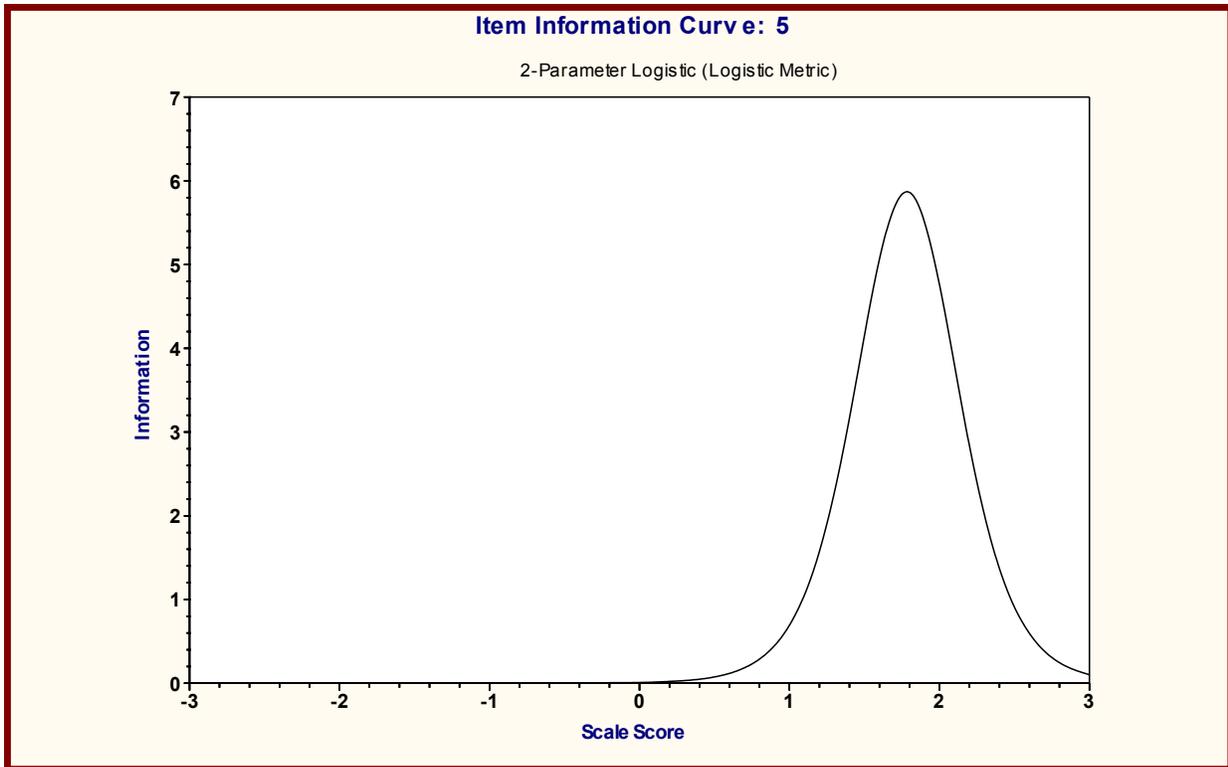
*Figure 3.* Item Information Function for “Lies to me”



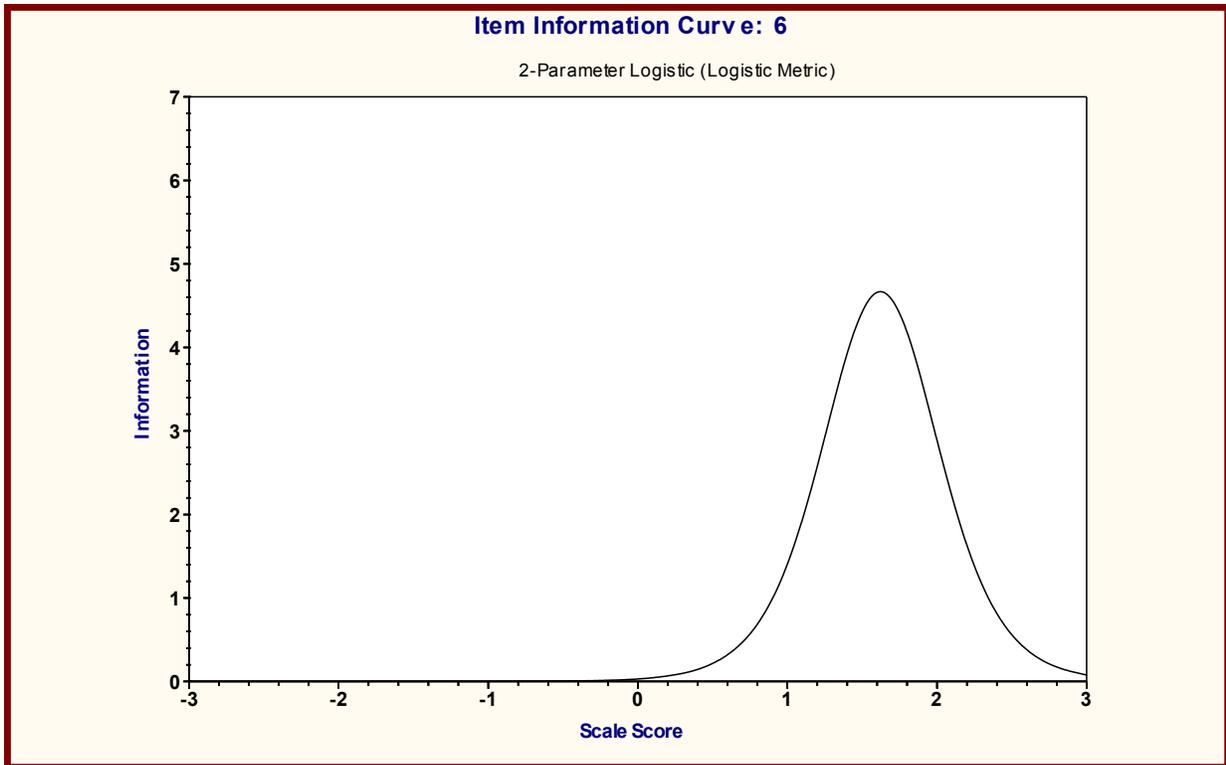
*Figure 4.* Item Information Function for “Would risk me to protect himself/herself in work matters”



*Figure 5.* Item Information Function for “Deliberately fuels conflict among employees”



*Figure 6.* Item Information Function for “Is evil”



*Figure 7.* Item Information Function for “Has it in for me”

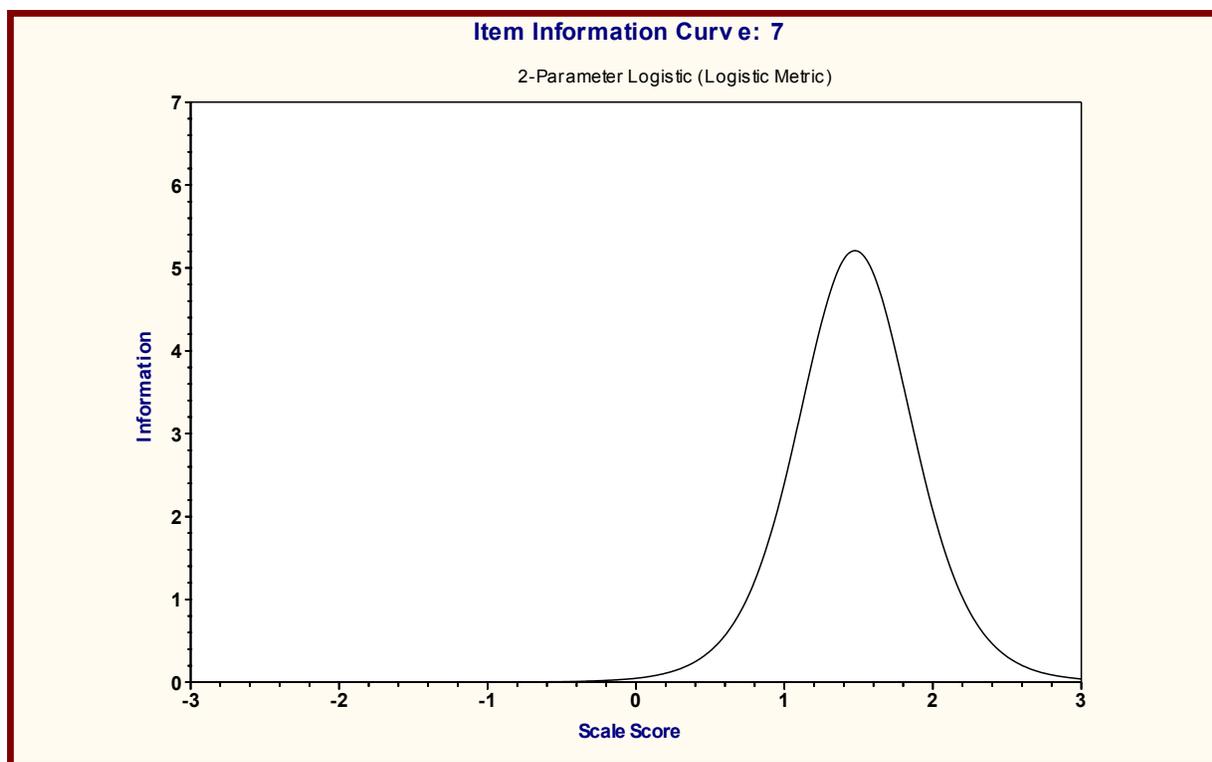


Figure 8. Item Information Function for “Is a hypocrite”

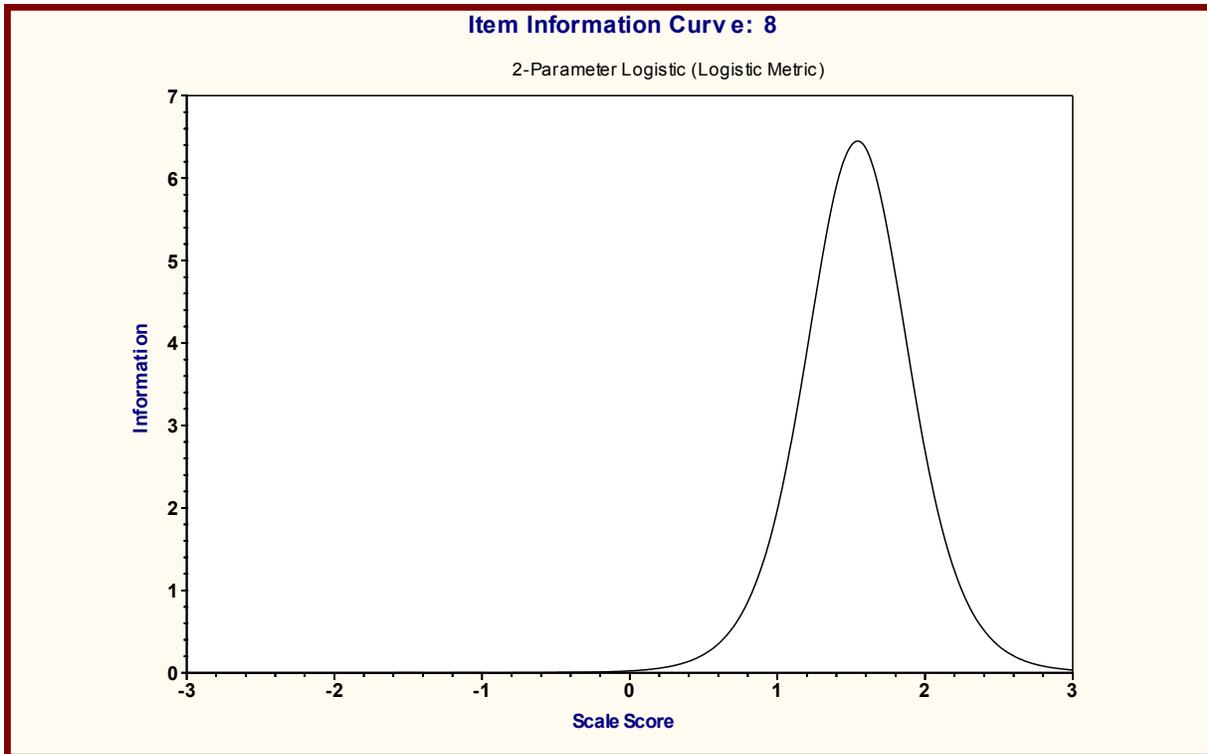


Figure 9. Item Information Function for “Would risk me to get back at someone else”

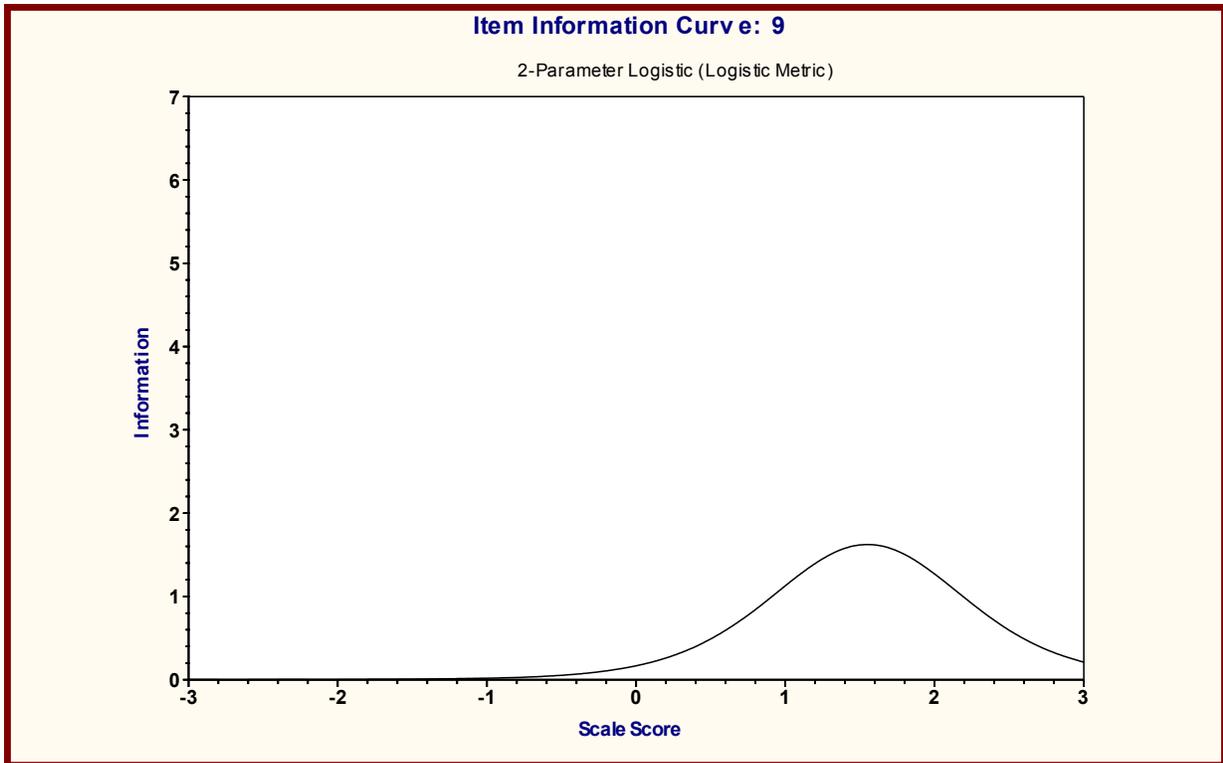


Figure 10. Item Information Function for “Would fire people just because (s)he doesn't like them if (s)he could get away with it”