A. **Introduction of the Instrument (e.g., title, authors, publication date)**

The title of the technical manual that I am examining is titled "The Job Training and Job Satisfaction Survey" published by Steven W. Schmidt in 2004.

B. **Trait and Behavioral Domain Definition**

The construct intended to be measured is job satisfaction. He used job training satisfaction to measure job satisfaction pointing to studies conducted by Traut (2000) and Mau and Kopischke (2001), which identified that job training is a specific factor in measuring job satisfaction. The author measures satisfaction towards job training by measuring the degree to which employee feels that their employers support the job training, the overall employee's feelings about job-specific or general training, degree of employee satisfaction with the job training, etc. The author also asked the respondent to rate the satisfaction with rewards and opportunities, satisfaction with supervisors, fringe benefits, nature of work, co-workers, and operating conditions.

C. **Intended Test-taker Population(s)**

The intended test-taker population is employees who are under permanent employment or contract employment.

D. **Intended and Unsupported Score Interpretations (Total and/or Subscale Scores)**

Based on page 16 of the technical manual, Schmidt gave the respondents 43 statements on job training and job satisfaction and asked the respondents to rate on a 6-Point Likert Scale. The point distributions for the 6-Point Likert Scale are as follow.

(a) Disagree very much – 1 point
(b) Disagree moderately – 2 points
(c) Disagree somewhat – 3 points
(d) Agree somewhat – 4 points
(e) Agree moderately – 5 points
(f) Agree very much – 6 points

The 43 questions attempt to measure job training and job satisfaction - 31 questions measure job satisfaction whereas 12 questions measure job training satisfaction. The 31 questions that measure job satisfaction surround the idea of 6 main constructs, namely rewards, supervision, benefits, operating procedures, co-workers, and the nature of work performed. The 12 questions that measure job training satisfaction are under the hood of 3 main constructs, namely organizational support, employee feelings, and satisfaction on training.

The author warns that for negatively coded items, those items shall be scored in a reverse manner as follows.

(a) Disagree very much – 6 points
(b) Disagree moderately – 5 points
(c) Disagree somewhat – 4 points
(d) Agree somewhat – 3 points
(e) Agree moderately – 2 points
(f) Agree very much – 1 point

If the respondents score 186 points on the job satisfaction survey, it means that they answered "agree very much" on each question and are overall very satisfied with their jobs. If the respondents score 72 points on the job training satisfaction survey, it means that they answered "agree very much" on each of the questions underneath it and are overall very satisfied with the job training. Based on the evidence portrayed by Schmidt that we will look at later, a high score on job satisfaction should also yield a high score on job training satisfaction.

E. Evidence for Validity of Score Interpretations for Each Intended Test-Taker Population

Evidence of a systematic process for development of the test specifications

Schmidt conducted a literature review to examine the relationship between job training and job satisfaction (Schmidt, 2004). He also confirms whether improving job training will improve job satisfaction through literature review (Traut et. al. 2000; Mau & Kopischke, 2001) and multiple other papers. To understand the demographics of his respondents, the authors also include questions such as the employee age, job status, formal education level, etc. On page 6 of the technical manual, we can see that the author also put some effort into defining the list of each of the training constructs. For example, for the construct, method of training, the author listed a few examples that include, instructor-led classroom, self-study, job-shadowing, etc.

Evidence of systematic item tryout and revision

Schmidt conducted a pilot study of this instrument with 118 people (Schmidt, 2004). He then uses factor analysis to remove nonessential questions based on the data collected from the pilot study. The main goal of the factor analysis as stated by the author is to ensure "concurrent-criterion validity" (Schmidt, 2004). The original job satisfaction survey was published by Spector in 1997 where he had an 8-factor solution consisting of rewards, promotion, and pay as the subscales of satisfaction (Spector, 1997). The author in this technical manual combined these criteria in Spector's job satisfaction survey into one category titled "satisfaction with opportunities and rewards" (Schmidt, 2004). Schmidt also removed the scale that measures satisfaction with communication from the existing Spector job satisfaction scale. The author conducts a Cronbach's alpha test, and it shows a value of 0.89 for the revised scale. The general rule is that Cronbach's alpha value above 0.70 is a good indication of acceptable consistency (Taber, 2018). The author states that the job training survey that consists of three major constructs has a Cronbach's alpha of 0.83, indicating a high consistency. The factors that correlate highly on the job training survey are employee satisfaction on personal and professional development, employee satisfaction with the amount of training received, and organizational support based on the factor loading. For the job satisfaction survey, the author conducted the

factor loading for a varimax six-factor solution and found that coworkers, operating conditions, fringe benefits, nature of work, supervision, and pay and promotion correlate highly.

Evidence that the items on each test form, as a set, are a representative sample from the behavioral domain

Schmidt used face validity to review each survey item for accuracy. This step is to ensure that the constructs have been precisely converted to items on this measurement tool. Schmidt also made reference to Spector's well-established work on job satisfaction scale while modifying only a few criteria (Spector, 1997). Schmidt then performs Cronbach's alpha test on the modifications done to ensure its consistency with the behavioural domain of job satisfaction.

Evidence of large, recent, representative norming samples for all intended test-taker populations

The author utilized a reasonable sample size, which is 118 people to validate the job training and job satisfaction instrument. The demographic questions consist of questions regarding gender, age, formal education level, job status and time in position. The author also shows that there is at least 100 employees answering each of the demographic questions out of the 118 respondents. Perhaps some of the respondents missed a few of the demographics questions unintentionally or they prefer not to disclose certain demographic information about themselves, so not all 118 respondents respond to all the demographic questions.

Evidence of relationships in expected directions (positive?, negative?) between scores from the instrument and criterion scores (e.g., scores from instruments measuring similar or unrelated traits, score on future outcomes)

Schmidt performed a comparison with the variables on the job satisfaction scale to that of the job training satisfaction scale (Schmidt, 2004). The author found a positive correlation, which enhances the construct validity.

Evidence that total scores and/or subscale scores are not substantially influenced by random measurement error (i.e., scores have adequate reliability)

The pilot study conducted by the author attempts to measure reliability. The majority of the items show high reliability, with alpha larger than 0.70 for 6 out of the 8 constructs. For the remaining 2 items, namely employee satisfaction on the training and operating conditions, though the alphas are less than 0.70, their values of 0.61 and 0.68 respectively still warrant marginally acceptable magnitude.

If particular cut score values are recommended for use in interpreting test scores, evidence that the standard-setting process was conducted systematically by a panel of qualified subject-matter experts

The author also employs three professors who are subject matter experts in the field of adult education to examine all education-relevant constructs and questions for validity. In my point of view, perhaps adding the population of non-experts but are employees of a particular company as the

evaluation panel will add further insights into the survey items as this non-expert panel is in a similar position compare to the population of the survey-takers.

F. Likelihood of Undesirable Consequences of Testing or Score Use

The undesirable consequence of the testing method is that we don't know which of the item requires extra cognitive effort compare to the other items as the author does not collect such data during the pilot study. There are several studies that indicate that extra scrutiny might affect data quality since the respondents would not answer the survey questions based on their gut feeling, but they will answer the questions based on social conformity. The idea is that in the real world, there is very little information available when we are trying to make a choice. Dijksterhuis et al. in 2006 discussed the "deliberation-without-attention effect"(Dijksterhuis et al., 2006). This term means that people will not make a good decision when they are more careful in their choices than they follow their gut feelings.

Further, the author doesn't account for outlier circumstances when the respondents are satisfied with the job but are not satisfied with the trainings done during the job. This study assume a positive correlation for all respondents in the sense that people who are satisfied with the training are satisfied with their jobs.

G. Overall Evaluation of Test Quality (Instrument and Intended Administration Conditions)

The test quality of this technical manual is perhaps considered decent though not perfect with quite a few evidences for validity of score interpretations. Still, there are a few defects that the author might want to take into account when developing a revised version of this instrument.

It would be ideal if additional data can be collected via a few more methods besides answering the survey. The authors could make the respondents think-aloud on their choices if resources allowed. That way, the authors can understand the choices made by the respondents. Further, the author might want to include an example explanation for each of those 6 choices for the respondents as a guideline. For example, when should the respondent choose "agree moderately" and "agree somewhat". The line between these two options seem to be unclear, and might left respondents pondering which option should they choose for a question such as "I view my education on-the-job as a continuous, lifelong endeavor".

H. Overall Evaluation of Validity of Intended Test Score Interpretations

Although there is several evidences used by the author, there is a few evidences that needs to improve. First of all, the authors used 3 professors who are experts in the field of adult education to evaluate the constructs. Besides utilizing the skills of experts for items validations, the author should also seek the opinions of non-experts and who are employees themselves to evaluate the items to see if those items would correlate to job training and job satisfaction. That way, the authors can have the viewpoint of the respondents who are similar to the survey takers population.

To identify if the respondents can answer all the questions smoothly, perhaps the author should consider recording the time it takes for the respondents to complete the survey. If the situation allows, it might be an excellent idea to record the time it takes for each individual to answer each particular question during the pilot testing phase. This way, the author can know that which question generally takes more time to answer, and the author can add more clarifications to the survey items that take longer to answer.

In terms of the 6-point Likert scale that the author used, it might be ideal if the author provide justifications to why this scale is used instead of a 5-point Likert scale or a 4-point Likert scale. The additional justifications on the scales will provide more insights to the scores that are obtained through the surveys, for example how would the choice of "disagree moderately – 5 points" differs from the choice of "disagree somewhat – 4 points".

I. **References**

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science (New York, N.Y.)*, *311*(5763), 1005–1007. https://doi.org/10.1126/science.1121629

Hanging on or Fading Out?: Job Satisfaction and the Long-Term Worker—Carol Ann Traut, Rick Larsen, Steven H. Feimer, 2000. (n.d.). Retrieved June 5, 2021, from https://journals.sagepub.com/doi/10.1177/009102600002900304

Mau, W., & Kopischke, A. (2001). Job search methods, job search outcomes, and job satisfaction of college graduates: A comparison of race and sex. *Journal of Employment Counseling*, *38*. https://doi.org/10.1002/j.2161-1920.2001.tb00496.x

Schmidt, S. W. (2004). The Job Training and Job Satisfaction Survey Technical Manual. In *Online Submission*. https://eric.ed.gov/?id=ED494451

Spector, P. E. (1997). *Job Satisfaction: Application, Assessment, Causes, and Consequences*. SAGE.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2