

Improving the Validity of Self-Ratings of Abilities: The Impact of Rating Format and Order

Ying Fong (echo.fong@voiceproject.com.au)

Voice Project, Department of Business
Macquarie University, Sydney NSW 2109 Australia

Joanne Earl (jearl@psy.unsw.edu.au)

School of Psychology
University of New South Wales, Sydney NSW 2052 Australia

Peter H. Langford (peter.langford@voiceproject.com.au)

Voice Project, Department of Business
Macquarie University, Sydney NSW 2109 Australia

Abstract

Previous research has found self-ratings of abilities to have low validity. Aiming to identify ways to improve self-ratings, this study investigated the impact of rating format (single-item global ratings versus multiple-item specific ratings) and order (global ratings before specific ratings versus global ratings after specific ratings) on the validity of self-rated abilities. A self-rating measure of communication skills was developed for this study for data collection from 178 medical students. The measure consisted of a global item assessing communication skills in general and eleven specific items, with each evaluating a particular dimension of communication skills. Students' self-ratings were correlated with the ratings from two groups of raters, namely communication tutors and clinical tutors, to establish validity. Results showed that students tended to provide more valid self-ratings on the specific items measuring different dimensions of communication skills than on the global item. We also found that the validity of students' self-ratings on the global item was improved when students rated themselves on the specific items prior to the global item. Theoretical and practical implications of the findings are discussed, and future research directions are suggested.

Introduction

Many career theorists have acknowledged that abilities (abilities, skills and competencies will be referred to collectively as "abilities" in this study) are one of the important psychological requirements that determine a person's career choice, suitability and development. The recognition of the importance of abilities has been translated into practice and reflected in the development of various vocational assessment tools used by educators and counselors in exploring career options with their clients. Rather than using objective measures of abilities, most career assessment tools have employed self-ratings of abilities as inputs to generate occupational possibilities. A few common examples

include the Self-Directed Search (SDS; Holland, Fritzsche & Powell, 1994), Campbell Interest and Skill Survey (CISS; Campbell, Hyne, & Nilsen, 1992), Career Decision-Making System (CDM; Harrington & O'Shea, 2000), and DISCOVER Computer-based System (American College Testing Service, 2004).

Although self-ratings of abilities have been commonly used over the years in the research and practice of career development, empirical evidence has provided little support for the validity of self-ratings. For example, Mabe and West (1982) found that "self-reports of ability evaluation frequently appear in the literature with little or no support for their accuracy" (p. 280). Mabe and West reviewed 55 studies in which self-ratings of ability were compared with measures of performance and reported a low mean correlation coefficient of .29. In another review of self-assessment studies in higher education, Falchikov and Boud (1989) reported the results of 44 studies in a variety of subject areas (e.g., medicine, law, engineering), and found the mean correlation between self-assessments and objective measures of ability or performance to be .39. Applying invalid self-ratings in career assessment may result in generation of less appropriate career options or a limited range of vocational alternatives. The current study aimed to identify ways to improve the validity of self-ratings.

In previous research on self-evaluation, the self-rating measures developed by the authors usually offered participants a single item on which to rate a particular ability, such as verbal reasoning (e.g., DeNisi & Shaw, 1977; Westbrook, Buck & Wynne, 1994). This also occurs in many commonly used career assessment tools. For example, the SDS requires individuals to self-rate each of twelve areas of abilities simply designated by name, e.g., "clerical ability" (Holland et al., 1994).

A single-item measure may suffice if the psychological construct being measured is sufficiently narrow or unambiguous to the respondent (Sackett &

Larson, 1990). Many abilities (e.g., mental toughness), however, are regarded as ambiguous because no objective indicators are available for these abilities or the criteria for evaluating them are not clear (Felson, 1981). Skills and performance are usually considered to be multi-dimensional rather than uni-dimensional (Harris & Schaubroeck, 1988; Levine, Flory & Ash, 1977). Many researchers regard self-estimates as self-beliefs or link self-estimates to self-concept or self-awareness (e.g., Gottfredson, 2002). Therefore, self-rated abilities may be viewed as complex psychological constructs rather than simple self-reported facts. Compared with single-item measures, multiple-item measures are more likely to generate valid ratings in the context of self-evaluation of ambiguous abilities.

With a single-item, individuals typically estimate an ability in a global fashion, without careful consideration about the specific dimensions of the ability. Atwater and Yammarino (1997) identified a number of factors affecting self-rating accuracy, one of which is the specificity of the dimensions of the ability to be rated. The authors suggest that greater agreement between self and others' ratings would be expected when individuals rate an ability that is clearly defined in terms of dimensions or aspects. The findings of other research provide some support for their proposition. For instance, Felson (1981) found that football players were more likely to over-rate themselves on ambiguous abilities (e.g., coordination) than on unambiguous abilities (e.g., strength). In another study, Wohlers and London (1989) found higher correlations between self- and co-worker ratings on managerial behaviours that were observable and clearly defined.

In the current study, Atwater and Yammarino's proposition was incorporated into the design of a self-rating measure. The ability subjected to self-evaluation was defined by a number of dimensions with observable behavioural examples to reduce its ambiguity. A self-rating measure that contained multiple specific items and a single global item was developed. Each of the specific items measured a particular dimension of the ability. Self-ratings on the specific items were compared with those on the global item. It was hypothesised that participants would provide more valid self-ratings when they rated their ability in terms of dimensions on multiple specific items than when they rated their ability on a single global item (Hypothesis 1).

Previous research suggests that self-ratings of abilities become more accurate as individuals gain more experience in self-evaluation (Kooker, 1974; Levine et al., 1977). This has some implications for the order of presentation of the global and specific items in the self-rating measure. Presenting the specific items before the global item in the questionnaire provides participants with the opportunity to consider each dimension of the

ability prior to rating the ability as a whole. This experience with the specific items might allow participants to gain a more thorough understanding of the ability being rated and to have some practice on self-evaluation before they rate themselves from an overall perspective, which may help improve the validity of self-ratings on the global item. Thus, it was expected that participants who completed the specific items prior to the global item would provide more valid self-ratings on the global item (Hypothesis 2).

Method

Participants

Participants of this study were 178 first-year medical students from the University of New South Wales (UNSW). The sample consisted of 83 males and 95 females, with a mean age of 19 years ($SD = 1.77$).

Seventeen communication tutors (4 males and 13 females) and 13 clinical tutors (5 males and 8 females) provided ratings for each participant as the criterion measures to establish the validity of students' self-ratings. All communication tutors held qualifications in psychology, social work or psychiatry, and had substantial experience in teaching counselling and communication skills. They were responsible for conducting a 3-hour workshop fortnightly at the campus. Each tutor supervised the same group of 10 to 12 students in every workshop. All clinical tutors were general physicians whose responsibility was to teach a half-day clinical tutorial fortnightly in the hospital. Each of the clinical tutors supervised the same group of 6 to 7 students in every tutorial.

When this study was conducted in the first semester, all tutors had completed at least 6 workshops or clinical tutorials with their students. Students were given many opportunities to demonstrate their communication skills through activities such as role-plays in the workshops and interviews with patients in the clinical tutorials, allowing their tutors to observe and feed back about their performance. Both types of tutors were responsible for grading the communication skills of each of their students in a formal assessment at the end of the semester.

Measures

Student Self-Rating Form A self-rating measure of communication skills was co-developed by the authors and the Faculty of Medicine, UNSW. Part I of the rating form contained 11 specific items that measured 11 dimensions of communication skills used in a patient interviewing context. Each item contained a number of behavioural examples. For example, the first item ("create rapport with patient") contained five

behavioural examples: “greet patient”, “introduce yourself”, “define your task”, “seek permission for interview”, and “express concern and willingness to help”. Students rated each item on a 5-point scale from 1 = “I have just started to develop this skill”, through 3 = “I am consolidating this skill, but still need some improvement”, to 5 = “I have mastered this skill”. The Cronbach’s alpha for the 11 items was .89. Part II of the rating form was a global item asking the students to rate their communication skills in general on the same 5-point scale.

To test the hypotheses, the presentation of Part I (specific items) and Part II (global item) was counterbalanced in the student rating forms.

Tutor Rating Form The rating forms used by the tutors were adapted from the student self-rating form, with minor modifications of wording. The communication tutor rating form contained 7 specific items corresponding with the first 7 items in the student self-rating form (items 8 to 11 were inappropriate for communication tutors to use as they measured clinical skills in addition to communication skills). The clinical tutor rating form contained 11 specific items identical to those in the student self-rating form. The Cronbach’s alphas for the specific items rated by communication tutors and clinical tutors were .90 and .97 respectively. Both tutor rating forms presented a global item on which to rate the student’s overall communication skills after the specific items.

Procedure

Students completed self-ratings after they had completed part of their communication training in the first semester. For the purpose of counterbalancing, each student was randomly assigned to one of two groups: Group A and Group B. Group A students ($N = 90$) completed the rating form in which the global item was presented prior to the specific items, while Group B students ($N = 88$) completed the rating form in which the specific items appeared before the global item.

The tutors completed ratings of their students independently in their own time. One hundred and seventy-eight communication tutor rating forms that matched all student self-rating forms and 73 clinical tutor rating forms that matched the corresponding student self-rating forms (i.e. 73 out of 178) were returned by the tutors.

Results

Differences between Group A and Group B Student Self-ratings

Independent samples t -tests were conducted to compare the self-rating scores of the two groups of students on

the specific items and the global item. Group A (global item first) students rated themselves significantly higher on the specific items ($M = 2.75$, $SD = .62$) than did Group B (global item last) students ($M = 2.42$, $SD = .58$), $t(176) = 3.66$, $p < .001$. Group A students ($M = 2.69$, $SD = .91$) also provided more favourable self-ratings on the global item than did Group B students ($M = 2.34$, $SD = .76$), $t(161) = 2.69$, $p < .01$.

Relations between Student Self-ratings and Tutor Ratings

In this study, the correlations between student self-ratings and tutor ratings were used as an indicator of the validity of self-ratings; that is, higher correlations were assumed to indicate more valid and accurate self-ratings. Table 1 presents the correlations between student self-ratings and tutor ratings on the specific and global items.

Table 1: Correlations between Student Self-ratings and Tutor Ratings.

	Correlation with communication tutor ratings (r)		Correlation with clinical tutor ratings (r)	
	Group A [#] ($N=90$)	Group B [#] ($N=88$)	Group A [#] ($N=30$)	Group B [#] ($N=43$)
Specific items	.20	.33**	.25	.32*
Global item	-.04	.26*	.15	.32*

[#] Group A = global item first; Group B = global item last

** $p < .01$, two-tailed.

* $p < .05$, two-tailed.

Correlations with Communication Tutor Ratings To test Hypothesis 1 that students would provide more valid self-ratings on the specific items than on the global item, the correlations between Group A student self-ratings and tutor ratings on the global item (i.e. when the global item was completed first) were compared with the correlations between Group B student self-ratings and tutor ratings on the specific items (i.e. when the specific items were completed first). As shown in Table 1, Group A student self-ratings had no significant correlation with communication tutor ratings ($r = -.04$), while Group B student self-ratings correlated significantly with communication tutor ratings ($r = .33$, $p < .01$). A significance test confirmed that the difference between these two correlation coefficients was significant at the .05 level. This means that student self-ratings and communication tutor ratings correlated significantly higher on the specific

items than on the global item, which provides support for Hypothesis 1.

Hypothesis 2 stating that students who rated the specific items prior to the global item would provide more valid self-ratings on the global item was also supported. A significance test showed that the correlation between student self-ratings and communication tutor ratings on the global item was significantly higher for Group B ($r = .26, p < .05$) than for Group A ($r = -.04$) students. This result demonstrates that Group B students, who were presented with the specific items prior to the global item, produced more valid self-ratings on the global item than did Group A students.

Correlations with Clinical Tutor Ratings Group B student self-ratings correlated significantly with clinical tutor ratings on both the specific items ($r = .32, p < .05$) and global item ($r = .32, p < .05$). Although a clear trend is evident, these correlations were not statistically significantly higher than those between Group A student self-ratings and clinical tutor ratings ($r = .15$ for the global item and $.25$ for specific items). Nevertheless, given the small sample size for the clinical tutors, and that the direction of the differences between the correlations (i.e. higher correlations for Group B than for Group A) is the same as the statistically significant differences found for the communication tutors, these results provide cautious support for our hypotheses.

Discussion

Using the ratings from the communication tutors as the criterion measure, the current study found (a) that self-ratings on specific items measuring different dimensions of communication skills were more valid than self-ratings on a global item measuring the skills as a whole, and (b) that self-ratings on the global item became more valid when the specific items were presented prior to the global item in the questionnaire.

Our first finding provides support for Atwater and Yammarino's (1997) proposition that self-rating accuracy can be enhanced by high specificity of the dimensions of the ability to be rated. The finding also builds on research that shows greater agreement between self and other ratings on more specifically defined abilities (Felson, 1981; Wohlers & London, 1989).

One explanation for the results may be that the global item is broad and ambiguous in defining communication skills, leading to more biased self-ratings and thus, attenuating the agreement between self and other ratings. In comparison, when communication skills are defined in terms of concrete and observable dimensions by using multiple specific items with clear

behavioural examples, participants tend to provide less biased self-ratings. This explanation is based on the argument that when traits are observable or verifiable (and therefore less ambiguous), individuals are less self-enhancing in evaluating themselves, resulting in more accurate self-ratings. (e.g., Dunning, Meyerowitz & Holzberg, 1989; Gosling, John, Craik & Robins, 1998; Klein & Buckingham, 2002). For example, Dunning et al. (1989) found that supplying participants with specific behavioural criteria for evaluating a trait reduced self-other rating biases. On the other hand, when the ability being assessed has high ambiguity, individuals may arrive at more favorable self-judgments, bringing about more biased self-ratings. Goethals, Messick and Allison (1991), for instance, found that individuals held greater self-other biases in evaluation of general rather than specific skills.

Another significant finding of this study was that participants generated more valid self-ratings on the global item when they completed the specific items first. With observable behavioural examples, the specific items illustrate various dimensions of communication skills. Presenting the specific items first may enable participants to gain a better understanding of the ability being rated and encourage them to make more careful self-judgments, resulting in more valid self-ratings on the global item.

The findings of this study have some important implications for the design of self-rating measures. First, single-item measures should be avoided as individuals are likely to produce more valid self-ratings on multiple specific items that measure different aspects of an ability. Developers of vocational assessment tools may consider providing a multiple-item dimensional measure for people to evaluate each of their abilities, skills or competencies, instead of using a single-item global measure. This helps generate more valid self-ratings as inputs in determining occupational options for career planning and decision making. In some research on self-evaluation, however, single-item measures of abilities may be preferred due to cost considerations or limitations of time and space on the questionnaire. In such cases, it is suggested that concrete behavioural examples be provided for each area of abilities to be rated.

Second, if a global item is included in self-rating measures, it should be presented after a number of specific items or scales, based on our finding that the order of presentation of the global and specific items affects the validity of self-ratings for the global item.

Future research may apply the design of the current study to investigate abilities more commonly used in career assessment and development, such as managerial skills. Researchers may also generalise the current design to investigate self-ratings in other contexts, such as 360 degree surveys used in organisations, as well as

other forms of ratings that involve evaluating other people or organisations, such as employee and customer surveys.

References

- American College Testing Service. (2004). *DISCOVER career guidance and information system*. Retrieved August 21, 2004, from <http://www.act.org/discover/index.html>
- Atwater, L.E., & Yammarino, F.J. (1997). Self-Other rating agreement: A review and model. *Research in Personnel and Human Resources Management, 15*, 121-174.
- Campbell, D. P., Hyne, S.A., & Nilsen, D.L. (1992). *Manual for the Campbell interest and skill survey*. Minneapolis, MN: National Computer Systems.
- DeNisi, A.S., & Shaw, J.B. (1977). Investigation of the uses of self-reports of abilities. *Journal of Applied Psychology, 62* (5), 641-644.
- Dunning, D.A, Meyerowitz, J.A., & Holzberg, A.D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082-1090.
- Falchikov, N. & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430.
- Felson, R.B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly, 44* (1), 64-69.
- Goethals, G.R., Messick, D.M., & Allison, S.T. (1991). The uniqueness bias: Studies of constructive social comparison. In J.M. Suls and T.A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 149-173). Hillsdale, NJ: Lawrence Erlbaum.
- Gosling, S.D., John, O.P., Craik, K.H., & Robins, R.W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74*, 1337-1349.
- Gottfredson, G.D. (2002). Interests, aspirations, self-estimates, and the self-directed search. *Journal of Career Assessment, 10* (2), 200-208.
- Harrington, T., & O'Shea, A. (2000). *The Harrington-O'Shea career decision-making system revised*. Circle Pines, MN: American Guidance Service.
- Harris, M.M., & Schaubroeck, J (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.
- Holland, J.L., Fritzsche, B.A., & Powell, A.B. (1994). *Self-Directed search technical manual*. Odessa, FL: Psychological Assessment Resources.
- Klein, W.M.P, & Buckingham, J.T. (2002). Self-Other biases in judgments of ambiguous performance and corresponding ability. *Psychologica Belgica, 42* (1-2), 43-64.
- Kooker, E. (1974). Changes in ability of graduate students in education to assess own test performance as related to their Miller Analogies scores. *Psychological Reports, 35*, 97-98.
- Levine, E., Flory, A., & Ash, R. (1977). Self-assessment in personnel selection. *Journal of Applied Psychology, 62*, 428-435.
- Mabe, P.A. & West, S.G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67* (3), 280-296.
- Sackett, P.R., & Larson, J.R., Jr. (1990). Research strategies and tactics in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp.419-489). Palo Alto, CA: Consulting Psychologists Press.
- Westbrook, B.W., Buck, R.W., & Wynne, D.C. (1994). Career maturity in adolescence: Reliability and validity of self-ratings of abilities by gender and ethnicity. *Journal of Career Assessment, 2* (2), 125-161.
- Wohlens, A., & London, M. (1989). Ratings of managerial characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology, 42*, 235-261.