Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: https://www.tandfonline.com/loi/imte20

Confidence weighted answer technique in a group of pediatric residents

Cheryl L. Hausman, Jeffrey C. Weiss, Janet S. Lawrence & Carter Zeleznik

To cite this article: Cheryl L. Hausman, Jeffrey C. Weiss, Janet S. Lawrence & Carter Zeleznik (1990) Confidence weighted answer technique in a group of pediatric residents, Medical Teacher, 12:2, 163-168, DOI: 10.3109/01421599009006692

To link to this article: https://doi.org/10.3109/01421599009006692

Confidence weighted answer technique in a group of pediatric residents

CHERYL L. HAUSMAN, JEFFREY C. WEISS, JANET S. LAWRENCE & CARTER ZELEZNIK, *Department of Pediatrics and Center for Research in Medical Education, Thomas Jefferson University School of Medicine, Philadelphia, PA, USA*

SUMMARY *Physicians who are excessively underconfident or overconfident about their knowledge may have impaired clinical judgement. Confidence weighting of multiple choice examinations asks test-takers to state how confident they are that the answers they selected are correct. This previously described method allows the examinee to receive 'overconfidence' and 'underconfidence' scores. This study was designed to test the hypothesis that these scores would correlate with faculty assessment of pediatric residents' confidence level as observed in the clinical setting. Thirty-three pediatric residents took an examination of general pediatric knowledge using confidence weighting method. Percentage of questions answered correctly ranged from 40 to 81%. There was an association between increasing overconfidence and* lower *examination scores* ($r = 0.58$; $p = 0.001$). *Increasing overconfidence was also associated with decreasing underconfidence* ($r = 0.38$; $p = 0.04$). *Five faculty members, the program director and the chief resident were asked to rate their perceptions of the residents' confidence on a Likert-type scale. The period of observation ranged from 9 months to 3 years. Linear regression demonstrated an association between underconfidence indices and observed confidence in the clinical setting* ($r = 0.39$; $p = 0.03$). *In addition, three of four residents who left the program had either over- or underconfidence indices greater than one standard deviation from the mean. These results indicate that the multiple choice examination with confidence weighting can predict residents who will be judged as underconfident by clinical preceptors. This finding is important in light of our impression that such house officers often have difficulties later during their residencies.*

Introduction

During medical school and residency, trainees are expected to learn a large number of facts. A physician must also develop a realistic awareness of how much he or she

knows about a particular subject. It has been postulated that "the amount of uncertainty one has about what one knows is a noncognitive attribute that plays an important role in decision-making" (Wigder & Garner, 1982). Presumably the uncertain, underconfident physician will tend to seek out more information (i.e. lab tests, X-rays) than is actually necessary for making the decision at hand. Such excess information-seeking behaviour could result in delays in initiating treatment and excessive medical expense. In contrast, the overconfident physician might be more likely to make a premature diagnosis because he or she is unaware of the limitations of the available information. This could lead to the administration of unnecessary treatments with the risk of medical complications.

Several methods have been designed to measure an examinee's realism about his/her level of knowledge (Rothman, 1969; Rippey, 1970; Echternacht, 1972; Palva & Korhoner, 1973; Rippey and Voytovich, 1982, 1983, 1985; Zeleznik *et al.*, 1988). These have all involved scoring adaptions to a multiple choice examination and generally require computer calculation to perform the grading. Smith & Rippey (1979) found that approximately 33% of medical students overvalued their knowledge (overconfident) while about 19% undervalued their knowledge (underconfident). More women were underconfident than men but there was no gender difference in overconfidence noted. Voytovich & Rippey (1982) evaluated knowledge realism with respect to scores on a test that required students to synthesize facts into a diagnosis list. They found that those students who overvalued their knowledge tended to make premature diagnoses. This correlation was found to be specific for the particular knowledge area involved. Zeleznik *et al.* (1988) compared overconfidence and underconfidence indices with third year clerkship examination grades. Highly overconfident students generally had lower scores on clerkship examinations, especially in pediatrics and internal medicine. There was no consistent pattern noted when exam scores were examined with relationship to the underconfidence index.

Zeleznik *et al.* (1988) also examined the underconfidence and overconfidence levels in students receiving high honors grades for their clerkship clinical evaluation. Although the results were not statistically significant, trends were noted and those with honors were less likely to be in the highly overconfident group and more likely to be in the highly underconfident group. Our review of the literature found no studies that deal with knowledge realism and clinical decision-making by interns and residents.

The purpose of this study was to detrmine if overconfidence and underconfidence levels shown by pediatric residents on an adapted multiple choice examination were associated with evaluations obtained on clinical rotations during the internship year. We were particularly interested in those house officers who were identified by the program director as having major problems during their training.

Methodology

A 100-item multiple choice examination was administered to 30 pediatric residents before the beginning of their internship year during PL1 orientation from 1985 to 1988. Three senior residents also took the examination during their PL3 year and their results have been included. The examination was a standard multiple choice test, the majority of questions were 'best answer' type and approximately one third were 'K-type' questions. The examination was made intentionally difficult, and the interns were not expected to know all the information. After each item on the test, the intern was asked to indicate how confident he/she was that the answer was correct. Five

options were given for levels of confidence: 100, 75, 50, 25 and 0%. The examination was scored much later in the year.

The tests were scored in two ways. A raw score was derived by calculating the percentage of items answered correctly. Over- and underconfidence indices were also generated for each intern, using the following method: a discrepancy score was calculated at each level of confidence. For example, if the intern said he/she was 100% confident that 12 test items were correct, all of these items should be answered correctly. If only eight answered correctly a discrepancy score of +4 was assigned at the 100% confidence level. If the intern stated that he/she was 50% confident on 20 items, yet answered 15 items correctly a discrepancy score of −5 was assigned. (At 50% confidence we would expect one half of the items to be correct.) All of the positive discrepancy scores were added together to create an overconfidence index, and all the negative scores were summed up to generate the underconfidence index. Therefore, each intern had an overconfidence and underconfidence index in addition to the standard raw score.

Towards the end of the academic year, six faculty members who evaluate resident performance while on the inpatient service, intensive care nursery, outpatient department and two community affiliates, were asked to rate the residents on a 7-point Likert-type scale with '1' being extremely underconfident and '7' being extremely overconfident. In addition, the program director and the chief resident were asked to rate the residents. All participants were blinded to the confidence indices derived from the examination.

The program director routinely collects performance information on each house officer after each rotation. This evaluation includes information on the resident's history and physical examination skills, interpersonal skills, fund of knowledge and use of other scources such as the laboratory, library and consultant opinion. The program director was asked to indicate from a list of the 33 residents which residents he considered to be 'problem' residents. The problem resident was defined as a house officer with serious interpersonal, organizational or knowledge problems.

Statistical analysis was performed using Statworks, a statistical package produced by Cricket software for the Macintosh system. Unpaired *t*-tests were used to compare means, linear regression was used for associations between continuous data and Chi square was used for associations between categorical data. Statistical significance was defined at $p>0.05$.

Results

Thirty-three residents completed the examination and a histogram of raw scores is found in Fig. 1. Overconfidence indices range from 0 to 39.7 (Fig. 2). Underconfidence indices ranged from 0 to 16.6 (Fig. 3). Faculty ratings of confidence ranged from 1 to 7 although not all faculty members used the entire scale. There was no correlation between either the raw score or the confidence indices with the American Board of Pediatrics Intraining Exam scores or our standard ratings of resident performance.

The overconfidence score was inversely associated with raw score ($r=0.58$; $p=0.001$), so that the higher the overconfidence index, the lower the raw score. The underconfidence index was positively associated with the raw score ($r=0.46$; $p=0.01$), indicating the higher underconfidence indices were associated with higher raw scores.

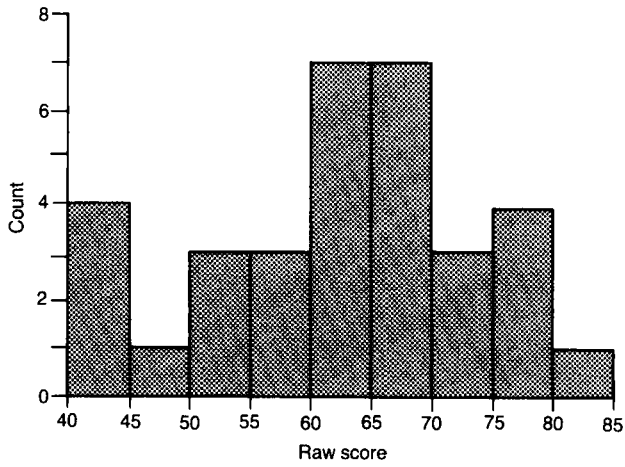

FIG. 1. Raw scores (mean=60, SD=10.4).

As expected, the overconfidence and underconfidence indices were inversely associated ($r=0.38$; $p=0.04$).

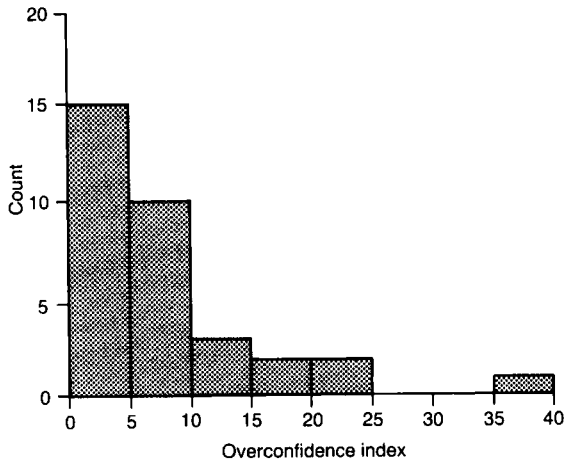

FIG. 2. Overconfidence indices (mean=8.5, SD=8.4).

Faculty predictions of confidence correlated with the residents' underconfidence indices ($r=0.39$; $p=0.03$). In fact, four of the faculty members had Pearson coefficients which were statistically significant. The overconfidence index, however, was not associated with faculty predictions ($r=0.09$; $p=0.64$). Pearson's coefficients for the association between overconfidence index and each individual faculty member's prediction of confidence were scattered over a wide range with the program director having the highest r value ($r=0.44$; $p=0.01$). Of the eight residents indentified as being problem residents, four had either over- or underconfidence indices >1 standard

deviation from the mean ($\chi^2 = 1.9$; $p = 0.2$) and three of four residents who left the program after the first year had usually high over or underconfidence indices. Of the nonproblem residents only eight of 25 tended to lie greater than one standard deviation from the mean.

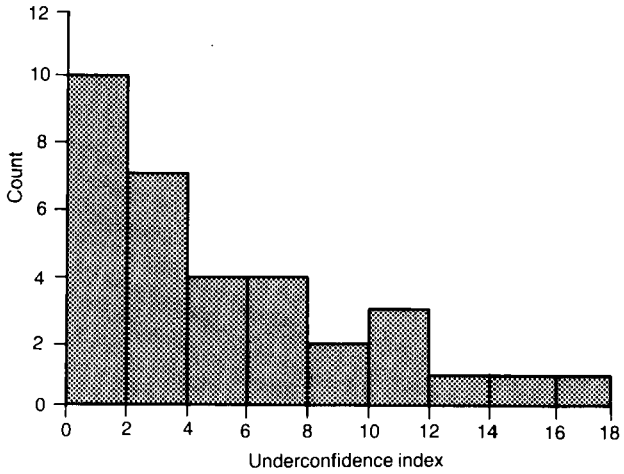

FIG. 3. Underconfidence indices (mean = 5.6, SD = 4.5).

Discussion

The underconfident resident may be defined as the resident who is never sure of what to do, and orders excessive laboratory tests and consultant opinions in order to be absolutely certain of a diagnosis. The overconfident resident, on the other hand, may jump to diagnostic conclusions (often incorrectly) underutilizing the laboratory or outside opinion. The underconfident resident may be preferable as his/her knowledge base is stronger as documented by the correlation with raw score. In addition, underconfidence may be a driving force for physicians who feel it important to continue to learn and read.

This study demonstrated that residents identified by the confidence examination to be underconfident were also often judged by their clinical supervisors as underconfident, thus demonstrating this scoring method to have empirical validity. Why faculty members were able to identify the underconfident resident more readily than the overconfident resident is not immediately apparent. The underconfident resident may be identified by tending more frequently to 'double check' his/her management decisions. In contrast the overconfident resident may ask few questions, choosing instead to act on his/her own decisions and may be noticed only if an error occurs.

To date, no single factor has been consistently shown to be of value in predicting resident performance. Our '*n*' was small and the incidence of 'problem' residents was also quite small. The trends found in this study, whereas not statistically significant due to these small numbers, may still be useful to program directors. The adapted multiple choice examination can be helpful to program directors by identifying a smaller pool of 'at risk' residents whose performance may require closer observation.

Whether or not 'confidence' is an inherent personality trait or is amenable to intervention is speculative. It may be possible to educate the very over- or underconfident resident to order tests and consultations appropriately in specific situations. These questions will require further study.

Conclusions

This report describes a test-scoring system designed to generate over- and underconfidence indices for each test-taker. There was an association between underconfidence and higher raw scores on the examination. Faculty were able to predict the underconfident residents but were unable (with the exception of the program director) to predict the overconfident resident. In addition, residents with unusually high over- to underconfidence indices were identified as 'problem residents' more often than expected. This scoring technique may help program directors to identify residents who are 'at risk' to have difficulties during their internship year.

REFERENCES

WIGDOR, A.K. & GARNER, W.R. (1982) *Ability Testing: uses, consequences and controversies* (Washington, DC, National Academy Press).

ZELEZNIK, C. *et al.* (1988) Students certainty during course test-taking and performance on clerkships and board exams, *Journal of Medical Education*, 63, pp. 881–891.

RIPPEY, R.M. (1970) A comparison of five different scoring functions for confidence tests, *Journal of Medical Education Measurement*, 7, pp. 165–170.

ECHTERNACHT, G.J. (1972) The use of confidence testing in objective tests, *Review of Education Research*, 42, pp. 217–236.

ROTHMAN, A.I. (1969) Confidence testing: an extension of multiple-choice testing, *British Journal of Medical Education*, 3, pp. 237–239.

PALVA, I.P. & KORHONER, V. (1973) Confidence testing as an improvement of multiple-choice examinations, *British Journal of Medical Education*, 7, pp. 179–181.

RIPPEY, R.M. & VOYTOVICH, A.E. (1983) Linking knowledge, realism and diagnostic reasoning by computer-assisted confidence testing, *Journal of Computer-Based Instruction*, 9, pp. 88–97.

RIPPEY, R.M. & VOYTOVICH, A.E. (1982) Adjusting confidence tests for realism, *Evaluation and the Health Professions*, 5, pp. 71–85.

RIPPEY, R.M. & VOYTOVICH, A.E. (1985) Anomalous responses on confidence scored tests, *Evaluation and the Health Professions*, 8, pp. 109–120.

SMITH, S.A. & RIPPEY, R.M. (1979) Confidence in test taking: a comparison of health profession students by sex and professional school, *Journal of the American Medical Women's Association*, 34, pp. 335–343.

VOYTOVICH, A.E. & RIPPEY, R.M. (1982) Knowledge, realism and diagnostic reasoning in a physical diagnosis course, *Journal of Medical Education*, 57, pp. 461–467.