General Article

IDENTIFYING COMPETENCIES WITH BEHAVIORAL-EVENT INTERVIEWS

By David C. McClelland

Boston University

Coding competencies from behavioral-event interviews according to the principles of a new approach to assessment (McClelland, 1973) produces assessments that are reliable and validly associated with success as an executive. These assessments are not influenced by length of protocol or by performance in the preceding year. Bias is not a problem if both the interviewer and the coder are blind to executive success. In contrast to a traditional psychometric approach based on regression analysis, an algorithm based on competency scores predicted managerial success and improved performance across a number of samples. This algorithm identified potential outstanding performers as individuals whose scores reached designated tipping points within clusters of substitutable competencies. Experts' judgments of competencies needed or shown by executives in various positions agreed only moderately with competencies shown to be important by the data from behavioral-event interviews.

Where does the competency-assessment movement stand today? Certainly it has come a long way from 1973, when I (McClelland, 1973) argued that competency assessment should be developed as an alternative to academic-type intelligence testing, which was failing to account for successful performance, especially in high-level executive positions. This view was strongly challenged by Barrett and Depinet (1991) on the grounds that intelligence tests were doing a good job and there was no evidence competency testing was any better. Nevertheless, evidence showed that Scholastic Assessment Test verbal and mathematical scores yielded zero or negative path coefficients for later occupational success in law, medicine, business, and teaching (Whitla, 1975) among students in an excellent liberal arts college, whereas competencies acquired in college did predict those same criteria of occupational success (Winter, McClelland, & Stewart, 1981). Furthermore, consulting companies, like McBer & Company, have completed dozens of studies showing that competency measures drawn from interviews predict success among high-level executives (Boyatzis, 1982), though many of these studies remain unpublished; in contrast, success at this level is seldom related to scores on academic intelligence tests. Also, Hogan, Hogan, and Roberts (1996) have demonstrated that personality variables unrelated to intelligence are related to occupational success, though rarely at the very high executive levels involved in the McBer studies.

David C. McClelland died on March 27, 1998. The final editing of this article was completed by David G. Winter, John Larrere, and Michele Nathan. Address correspondence to David G. Winter, Department of Psychology, University of Michigan, 525 East University Ave., Ann Arbor, MI 48109-1109; e-mail: dgwinter@umich.edu.

Although researchers have attempted to develop competency tests for wider use on large samples (e.g., McClelland, 1994; Winter et al., 1981), such tests have not seen wide acceptance so far. However, the interview method of assessing competencies has been widely accepted and has led to a significant amount of research, which can be used to test its strengths and weaknesses. Furthermore, of all methods, this one comes closest to implementing the four propositions I put forward (McClelland, 1973) as the basis for a new approach that would be favorable to good competency assessment: (a) It should start with exploration of particularly operant measures of thought and action associated with the criterion—in this case, greater success in high-level executive and managerial positions; (b) it should assess clusters of criteria of success in clusters of important life outcomes (e.g., occupations, health, family and social life, education); (c) the competencies assessed should be defined and described in ways that reflect important life changes or learning; and (d) how to improve on the competencies should be studied, and made explicit and public. It can readily be seen how this approach differs from those in which secrecy is required to protect the integrity of the test scores (as in IQ testing), or in which someone is simply rated on a competency, without getting information about the behaviors on which the rating was based—information that could lead to improved performance.

THE BEHAVIORAL EVENT INTERVIEW

The Behavioral Event Interview (BEI) is an adaptation of the critical-incident interview originally developed by Flanagan (1954) and later elaborated by Dailey (1971) and, especially as employed in the present study, Boyatzis (1982). The

BEI was designed as the most flexible way to discover differences between two types of job incumbents: those who have been nominated by knowledgeable judges as outstanding (O) and those who have been nominated less often or not at all (referred to as typical, T). This approach is used because people agree more readily on who is outstanding than on what makes them outstanding, and because having judges rate characteristics supposedly related to success (rather than rating actually successful people) might result in a biased criterion. Generally, the O group is in the top 5% to 10% of the executives, and the T group includes the next 11% to 25% of the executives.

In specially designed interviews, the O and T groups describe, in their own words, what they said, thought, felt, and did in six episodes—three positive and three negative—at work. Standardized procedures have been established for ensuring interviews are comparable (Spencer & Spencer, 1993). The interviews are recorded, typed up, and coded for various characteristics. What is unique about this approach is that the competencies are defined and then redefined to improve the degree to which they distinguish between O and T performers on particular jobs. If a competency is found to differentiate these two groups across samples of executives, it becomes part of a standardized dictionary of competencies.

When BEI transcripts are used in the fashion just described, they have an exploratory purpose for constructing competency models. That is, the comparison of O and T groups allows the development and definition of competencies that differentiate the two groups. In addition, BEI transcripts can be used as an assessment tool. For example, as I discuss later, transcripts of applicants for executive positions can be scored for previously defined competencies to help identify which applicants have the greatest potential for success.

Currently, competencies are coded both for frequency of occurrence in the interview (Boyatzis, 1982) and for the level of complexity or scope at which they are displayed (Spencer & Spencer, 1993). For example, a competency labeled Impact and Influence is scored whenever the person describes using deliberate influence strategies or tactics. It is scored at a low level for “Takes a single action to persuade,” at a higher level for “Takes multiple (two or more) actions to persuade,” and at a still higher level for “Uses complex influence strategies, assembles political coalitions, etc.”

Average interjudge agreement for well-trained coders across a variety of competencies is in the range of 74% to 80% for maximum level and frequency of occurrence (Boyatzis, 1982; Nygren & Ukeritis, 1993). Motowidlo et al. (1992) reported high stability of their competency ratings across two interviews for lower level employees, and Boyatzis, Cowen, Kolb, and Associates (1995) reported significant stability for ratings of some competencies across 2 years for business-school students.

Within the limits set by the standard procedure, interview length does not affect competency scores. For a sample of 251 interviews, the scores for 22 competencies had an average correlation with interview length, after Z transformations, of $r = .06$ ($SD = .07$) for the frequency measure and $r = .04$ ($SD = .05$) for the maximum-level measure.

To eliminate the possibility of bias, neither the people interviewed, the interviewers, nor the coders know who has been nominated as outstanding or typical.

OVERALL TESTS OF VALIDITY DRAWN FROM PREVIOUS COMPETENCY RESEARCH

The evidence of validity lies in whether a high score on a competency measure is more often associated with successful performance across a variety of positions than one would expect by chance. Table 1 lists 12 competencies (from the latest revision of Spencer and Spencer’s, 1993, dictionary) that most often emerge as validated differences between O and T performers. The table summarizes the percentage of times each has differentiated between O and T samples of executives or professionals. The data come from more than 30 different organizations and many different types of executive positions, including managers, salespeople, mining geologists, bankers, restaurant managers, and heads of long- and short-term health care units. The number of persons in each comparison varied from 8 to 78, with a median of 12. In most competency-assessment projects, more O performers than T performers are interviewed, in a ratio of 3:2, because there are more ways of being outstanding than typical. The comparisons summarized in

Table 1. *Validity frequencies for core Behavioral Event Interview competencies*

Competency	Measure: Level	Measure: Frequency
Achievement Orientation	42**	66***
Analytical Thinking	42**	41***
Conceptual (Inductive) Thinking	31+	47***
Developing Others	38*	47***
Flexibility	35*	26*
Impact and Influence	46**	67***
Information Seeking	35*	41***
Initiative	31+	48***
Interpersonal Understanding	31+	32*
Organizational Awareness	31+	27*
Self-Confidence	46**	58***
Team Leadership	42**	30**

Note. The numbers in the table are the percentage of comparisons in which each competency differentiated the outstanding (O) and typical (T) groups in the expected direction, $p < .10$. For level, there were 26 comparisons for each competency, and for frequency, there were 64. The expected direction was a greater level of competency or a higher frequency for the O group than the T group.
+$p < .10$. *$p < .05$. **$p < .02$. ***$p < .001$.

Table 1 are based on competency-assessment data from 238 persons classified as outstanding and 225 considered typical.

For each competency, information was available for a sample of 64 O-versus-T comparisons of frequency and 26 O-versus-T comparisons of maximum level. The percentages of positive validities are well above the 10% chance level for most of the competencies in Table 1.

Although these data testify to the overall validity of these competencies, they also show that any one competency is not valid in all situations. The table lumps together very different occupations, and competencies that are valid for predicting success as a banker, for example, are not likely to be valid for a mining geologist. When enough cases are available, it will become possible to compare O and T groups within general types of executive positions, such as sales, finance, and personnel positions, so that one might be able to identify competencies that cross-validate within these categories.

Furthermore, even executives of the same type may work in significantly different organizational climates. For example, the need for Achievement (or *n* Achievement, as measured with the Thematic Apperception Test, or TAT; see McClelland, Atkinson, Clark, & Lowell, 1953) generally relates to entrepreneurial behavior and success (see McClelland, 1961). Thus, Andrews (1967) long ago showed that in one large organization that emphasized entrepreneurship within decentralized units, higher *n* Achievement appropriately was associated with managerial success. In another large company, Andrews found that *n* Achievement was not associated with success, but an examination of climate showed organizational clarity was significantly lower in the second company. Thus, managers with high *n* Achievement in the second company were rendered ineffective because they did not know clearly what they were supposed to be doing.

In a particular comparison of O and T groups, the focus is seldom on a single competency, but rather on the pattern of competencies that show significant differences. In the 26 organizations where competency levels were studied (results summarized in the left column of Table 1), O and T groups always showed significant differences on more than one competency: The range was from 3 to 14 significant differences, with a median of 8; in 85% of the organizations, there were significant differences in level on six or more competencies ($p < .10$ in the predicted direction). A study by Nygren and Ukeritis (1993) illustrates how a series of different individual competencies can also be worked into an overall meaningful pattern that explains in an integrated fashion how the more successful executives approached their job in the context of the particular institution studied.

Competency scores may also be interpreted in relation to computerized norms of various types, although such norms—based on aggregated data from many positions and organizations—are not of much value because they have no overall validity. Rather, a company usually wants to know whether the particular competency scores it may consider using for assessment purposes are related to job success in its own organization. For this purpose, a company needs to use "norms" based on managers having the same role or function, in successful companies of the same type.

The BEI competencies that relate significantly to managerial success differ from case to case, but does that mean no generalizations can be made as to what combinations of competencies will predict success across samples even within the same company? In fact, some generalizations are possible, because competencies within certain broad categories may substitute for each other. For example, unusual individual initiative can be reflected in either trying to do things better (Achievement Orientation), planning and thinking ahead (Initiative), or seeing things in a new light (Conceptual Thinking). Similarly, the ability to work well organizationally can be represented by influencing other people (Impact and Influence), by understanding organizational politics (Organizational Awareness), or by showing Team Leadership.

Thirteen studies of managers were examined to see whether the O group satisfied the following algorithm: mean frequency or maximum-level score significantly higher than that of the T managers (a) on at least one of the initiative and one of the organizational competencies and (b) on a total of five competencies drawn from the list in Table 1. The O groups in 11 (85%) of the studies satisfied this algorithm, compared with only 1 out of 8 (13%) studies of individual contributors, that is, technical and professional personnel such as geologists, consultants, and insurance raters ($p < .01$ for the difference in proportions). Thus, competency algorithms that are associated with success in various types of executive positions can be found by using the principle of substitutability; that is, a variety of different but functionally equivalent alternative predictor variables may be related to an outcome criterion. To some extent, therefore, different competencies can substitute for each other.

PERFORMANCE-BASED STUDIES OF VALIDITY IN ONE LARGE MULTINATIONAL CORPORATION

Moving From Nominations to Hard Performance Data About Executive Success

Testing validity by comparing BEI competencies in nominated O and T samples of job incumbents is a method with inherent weaknesses. Such comparisons may capitalize on chance differences. They rest on people's opinions of who are outstanding performers, not on measures of actual performance. And they may reflect past performance rather than predict future performance. In addition, such samples tend to be small, so why not study more people by using performance data that are already being collected for other purposes?

Fortunately, data on executive performance were available from "Tastyfood" (a pseudonym for a large food and beverage company). These data, which spanned several years, made possible a stronger and more objective test of validity. At the end of each year, each executive received a bonus based on the extent to which the executive had achieved performance goals agreed to at the start of the year. Because average bonuses awarded in various units of the company varied from year to year depending on unit profitability, the bonus scores were standardized for each unit in each year (M = 50, SD = 10; median N = 112, range: 53–242). Outstanding performance was defined as receiving a standardized bonus score of 55 or better (i.e., receiving a bonus in the top third of amounts awarded).

Combining Competency Measures

Eleven BEI-based competencies and one—a combination of moderately high achievement motivation and affiliation motivation—drawn from a six-picture modified TAT showed significant mean differences between O and T executives in an initial study. Traditional psychometric procedures, based on additive regression models, did not yield stable results. In one regression analysis, the previously validated competencies yielded a multiple R of .52 (p < .01) with the bonus criterion; however, in a second sample of 42 similar executives, the same regression formula failed to predict the same criterion (R = .24, n.s.). In this latter sample, a regression model based on different competencies did predict the criterion (R = .58, p < .01), but that model yielded an R of only .08 when applied back to the original sample.

Furthermore, in this and other samples, many of the relationships between success and frequency or level of competency were not linear and not well described by correlation coefficients. Instead they were characterized by what have been called “tipping points” (Gladwell, 1996). Sociologists have often observed that changes in a societal variable make little difference until they reach a certain level. For example, delinquencies and teenage pregnancies increase sharply in ghetto neighborhoods only when the number of middle-class people living there decreases to a critical low level (Crane, 1991). If the number of police in a neighborhood reaches a certain critical point, crime drops markedly. Increases in police presence below (or above) that level make little difference.

Something similar characterizes the relationship of competency frequencies and levels to success as an executive. For example, Figure 1 shows that for Impact and Influence, the T group is more likely than the O group to have a frequency score anywhere from 0 to 7; the O group is more numerous than the T group only when the frequency score reaches 8 to 10; further, this O-versus-T difference does not change at higher frequency scores, above 10. So it would be a misrepresentation of the relationship to describe it in terms of a linear correlation coefficient. For the data graphed in Figure 1, for example, the biserial r is .22, p < .10, between O versus T status and frequency of the competency Impact and Influence, but this statistic understates the significance of the relationship (55% of the O executives vs. 20% of the T executives had frequencies of 8 or more, p < .001) and misrepresents its nature for frequencies below 8.

Because of such patterns in the data, critical frequencies or levels that differentiated best between the O and T executive groups were established for the 12 valid competencies in this comparison. Then a competency-qualification algorithm was developed. It required that tipping points be achieved for at least 1 of the 3 individual-initiative competencies, 1 of the organizational competencies, and 6 of the 12 valid competencies overall. (Reaching the tipping point for more than 6

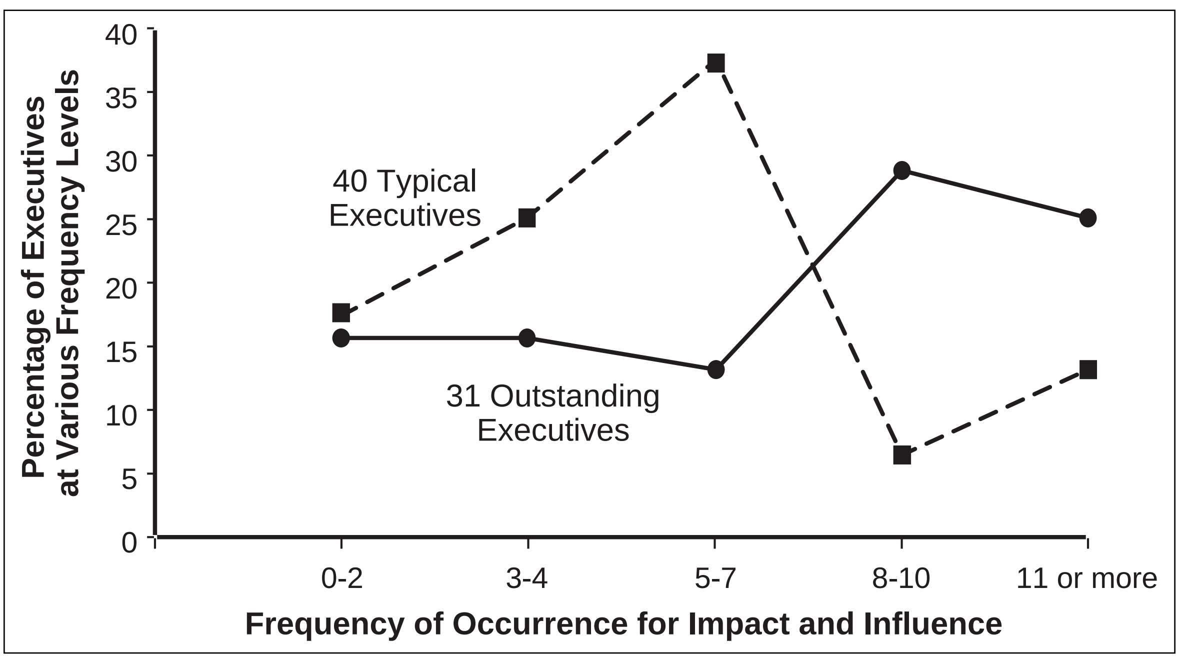

Fig. 1. Percentages of outstanding and typical executives showing different frequencies of the competency Impact and Influence.

competencies did not substantially increase the likelihood of being in the O category.)

The critical question was whether the competency algorithm predicted actual performance in additional samples. Table 2 presents findings for two U.S. and two overseas samples. The criterion of performance was standardized bonus awarded in the U.S. samples and nomination in the overseas samples. The relationships in the table are shown in terms of percentages of those executives predicted to be high and low performers, respectively, who actually turned out to be top performers. Such analyses are easy to understand and are evaluated by chi-square tests because the prediction variables are not related to outcomes in a linear fashion, but instead involve tipping points. In the initial U.S. sample, 100% of the executives who attained the competency-qualification algorithm were in the top third of bonuses awarded, as compared with only 29% of those who did not attain it, $\chi^2(1, N = 29) = 14.45$, $p < .01$ (tetrachoric $r = .90$, $p < .01$; 83% correctly predicted overall). This result is not surprising because the algorithm was derived from this sample. But applying the algorithm to a new sample of executives for the same bonus year also showed it was strongly related to actual performance, $\chi^2(1, N = 42) = 8.58$, $p < .01$ (tetrachoric $r = .65$, $p < .01$; 74% correctly predicted overall). Finally, in still a third sample of executives (not shown in Table 2), for the 1994 bonus year, the results were very similar. Of the 14 executives achieving the competency algorithm, 71% received bonuses in the top third of their distributions; in comparison, only 27% of the 11 executives who did not satisfy the competency algorithm received such large bonuses, $\chi^2(1, N = 25) = 4.81$, $p < .05$ (tetrachoric $r = .64$, $p < .02$). Furthermore, as Table 2 shows, the algorithm worked very well in predicting successful executive performance overseas when competency criteria that apply only overseas were included (for Europe and Asia samples combined, $\chi^2[1, N = 35] = 17.93$, $p < .001$; tetrachoric $r = .90$, $p < .01$; 86% correctly predicted overall).

Effects of Previous Year's Performance on Competency Scores

Competency scores might be influenced by how well people performed in the previous year; that is, during the interviews, the O group might describe more successful behavioral events than the T group, which would lead to higher BEI competency scores for the O group. If this were so, one would expect that in the data for Tastyfood, the average 1994 bonus scores would be higher for executives achieving the competency algorithm in 1995 than for those who were less highly qualified in 1995. The average 1994 bonus score was 56.12 ($SD = 7.63$) for 20 executives classified as highly qualified in 1995 and was 50.59 ($SD = 10.24$) for 8 executives not classified as highly qualified in 1995, an insignificant difference.

An additional analysis showed that of the 11 executives who were considered highly qualified by the competency algorithm in 1992 or 1993, 55% showed an outstanding performance in the previous year (i.e., received a high bonus). Of the 12 executives who appeared less qualified in 1992 or 1993, 50% had received a high rather than a low bonus in the previous year. Performance in a given year does not appear to affect whether an individual satisfies the competency algorithm the next year.

Do Competency Scores Predict Performance in the Following Year?

Table 3 addresses the issue of whether competency scores predict performance in the following year. It shows that competency, as assessed at Tastyfood using the algorithm in 1992,

Table 2. *Association of executive success with a competency algorithm based on Behavioral Event Interviews*

	U.S. samples[a]				Overseas samples[b]			
	Initial		New		Europe		Asia	
Performance predicted by competency algorithm	*N*	Top performer (%)	*N*	Top performer (%)	*N*	Top performer (%)	*N*	Top performer (%)
High[c]	12	100	17	65	11	82	7	86
Lower	17	29	25	20	8	13	9	11

[a]The measure of performance was bonus received in 1993, standardized by company unit ($M = 50$, $SD = 10$). An executive was classified as a high, or top, performer if his or her standardized bonus score was 55 or higher, representing the top third of the bonus distribution.
[b]The measure of performance was being nominated as an outstanding performer for 1993.
[c]To be classified as highly qualified by the competency algorithm, an executive had to score at or above the frequency that differentiated outstanding from typical performers on at least six (U.S.) or seven (Europe and Asia) of a specified list of competencies.

Table 3. *Mean standardized bonus scores in 1993 for executives assessed in the same or previous year as highly versus less qualified*

Assessment by competency algorithm	Assessed in 1992 n	Assessed in 1992 Mean bonus score	Assessed in 1992 SD	Assessed in 1993–1994 n	Assessed in 1993–1994 Mean bonus score	Assessed in 1993–1994 SD	Difference between means
Highly qualified	13	60.65	4.09	17	54.64	6.89	6.01**
Less qualified	18	45.71	12.02	26	47.56	9.65	–1.85
Difference between means		14.94***			7.08**		7.86+

Note. Bonus scores were standardized across each executive's individual unit (M = 50, SD = 10).
+p < .06. **p < .01. ***p < .001.

very strongly predicts performance as reflected in standardized bonuses awarded in 1993. In fact, the 1992 assessments predict the 1993 bonus scores even more strongly than competency assessments during or after the 1993 bonus year. Why?

An important purpose of the competency-assessment program was to provide participants with information on the nature of the competencies that were associated with successful performance in the positions they held. Accordingly, each executive was given a report that showed just what the tipping points were for each competency associated with success in the position and where his or her competency scores stood relative to these criteria. In individual sessions, executives were counseled about how to set goals and work for improvement in the relevant competencies.

Therefore, the results in Table 3 may be due to the fact that executives who received competency feedback in time to improve their performance did better than those who received feedback at a later time. Other research has shown that receiving feedback on competencies combined with setting goals for changing them improved competencies as measured 2 years later (Boyatzis et al., 1995).

Does Feedback on Competencies Improve Subsequent Performance?

The data from Tastyfood were examined to determine whether the timing of feedback affected improvement in performance from one year to the next. For each executive, the change from the bonus awarded one year to the bonus awarded the following year was determined. These bonus-change scores were adjusted by the overall regression of bonus change on initial bonus level so that it could be determined whether each executive's second bonus was better than expected based on his or her starting standardized bonus score. The findings are summarized in Figure 2. For all executives, whether highly qualified or not, 62% of those who received competency feedback in the year before the second bonus received a larger bonus than expected, as compared with only 39% of those who received the competency feedback in the year of the second bonus, $\chi^2(1, N = 90) = 4.85$, $p < .03$.

This analysis confirmed the hypothesis that executives show larger improvements in performance (as reflected in bonuses awarded) if they have a year to benefit from competency feedback than if the feedback is provided at a later time.

Executive Turnover as a Measure of Validity of the Competency Algorithm

A major reason why Tastyfood undertook the BEI-based competency program was to reduce turnover, which was very expensive. According to the company's figures, 17 of the 35 executives (49%) they had hired in the usual way in 1992 at roughly the vice-presidential level had left the company by the end of 1994, either voluntarily or through being managed out. Each lost executive at this high level cost the company about $250,000, so the poor selection system cost the company more than $4 million over this time period (see Fig. 3).

In comparison, 10 executives at this level were hired in 1993 guided by the BEI-based competency algorithm already described. By the end of 1995, only 1 of these executives had left the company (10% turnover). An additional 22 executives were hired in the first half of 1994, and if the 1993 and 1994 hires are added together, only 2 of 32 persons from this larger sample had left the company by March 1996, yielding a turnover rate of 6.3%. The decreased turnover rate ($p < .01$) suggests the company may have saved up to $3.5 million by using a competency-based hiring and feedback system in 1993 and the first half of 1994. It is not surprising that the competency-based program decreased turnover because it explained to new hires what characteristics promoted success on the job and because it produced a better job-person match (i.e., persons with more of the competencies needed for success were hired).

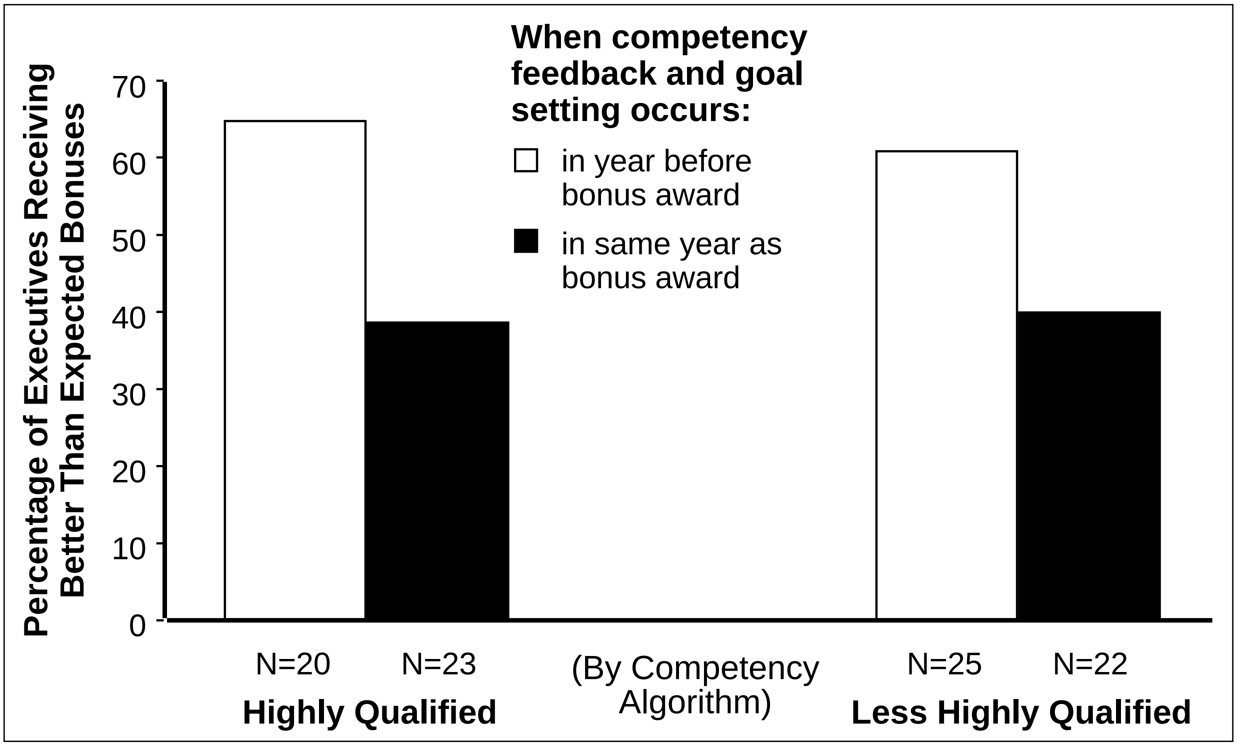

Fig. 2. Percentages of executives receiving better than expected bonuses as a function of their assessment using the competency algorithm. Whether a bonus was better than expected was determined by an overall regression of bonus change (change from initial bonus to second bonus, the following year) on initial bonus level. Results are shown separately for executives who received feedback and help in goal setting the year before the second bonus was awarded and those who received feedback in the same year as the second bonus was awarded.

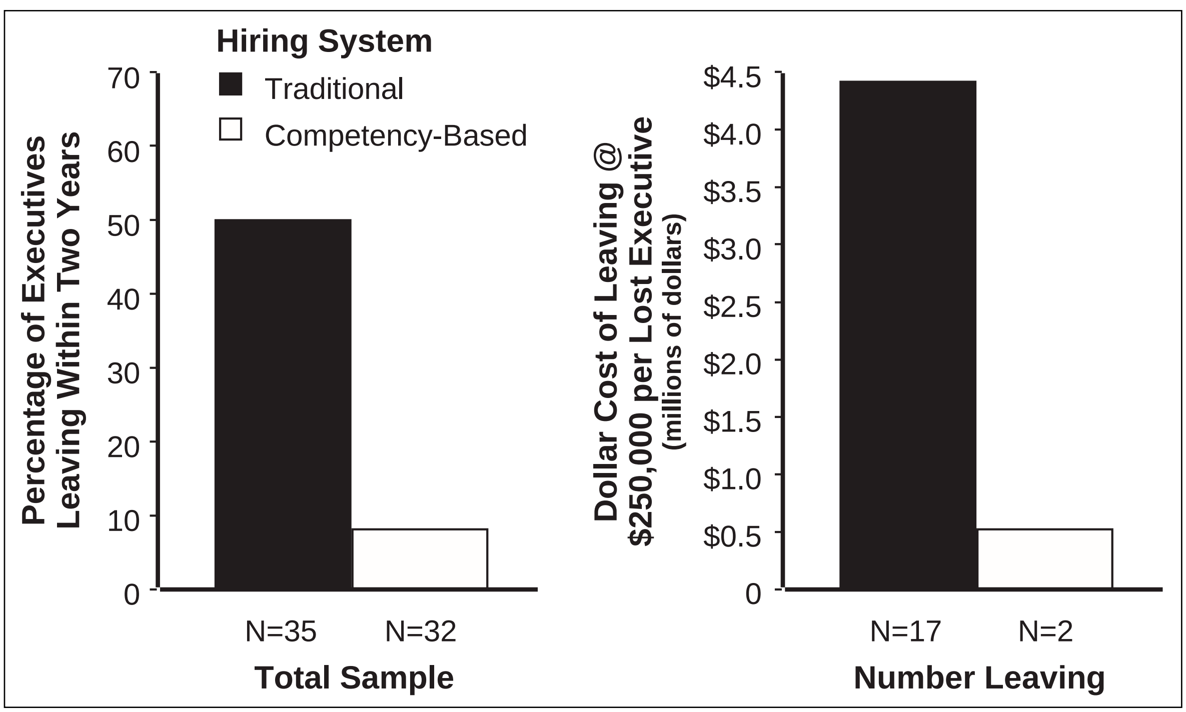

Fig. 3. Executive turnover rates and costs for the traditional and competency-based hiring systems.

Did the Competency-Based Algorithm Predict Performance in the Year After Executives Were Hired?

Whether the competency algorithm should be used for selection was examined by analyzing the data for 44 newly hired executives whose standardized bonus scores for 1995 were based on a full year's employment. Of the 17 executives who passed the competency algorithm when they were hired, 47% received standardized bonus scores of 55 or higher, compared with 22% of the 27 executives who did not pass the competency algorithm ($p < .05$ in the expected direction). Examination of the results suggested that this difference in proportions, although significant, was reduced because the competency tipping points were lower for the new hires (who came from varied backgrounds) than for long-term employees of Tastyfood.

Can Competency Ratings Substitute for Interviews?

The research reviewed so far involved competencies assessed through interviews and a competency model or algorithm based on actual outcome measures (nomination or objective performance data). However, an alternative method of constructing competency models and assessing competencies involves extensive use of ratings by "experts." Thus, expert panels, focus groups, or job incumbents may judge what competencies are needed for successful performance in a particular position. Then individuals may be rated on the degree to which they show the competencies judged to be required, and the extent of job-person match can be used to select or promote people (Caldwell & O'Reilly, 1990). This procedure is inexpensive and easy to carry out and has face validity. It is often used in a variety of settings, as reflected by many articles in the journal *Competency.* But is this ratings-based procedure accurate?

In a typical case of competency-model construction, 10 competencies (frequencies or levels) were identified from BEIs as distinguishing nominated O and T sales managers. Experts correctly identified 7 of these as important and missed 3; they also judged to be important 2 competencies that the BEI procedures did not identify as characterizing outstanding performers on the job. The overall agreement between the experts and the BEI results was 74%. However, the judges correctly identified only three of the critical levels for the 7 competencies that best distinguished between the O and T performers in the interviews.

In another comparison, 40 executives were rated by judges who knew them well, as either superiors, peers, or subordinates. The judges rated four to six items for each of 12 competencies. (The items described behaviors drawn from the actual competency definitions used to code BEIs.) Individuals were classified as "high" on a given competency if they scored at least at the tipping point for it in their BEIs or if their average score for the component items was above the median average score of all executives rated on that competency. For 43% of the 7 BEI competencies that significantly predicted a high bonus score, the judges' ratings and BEI scoring agreed (3 out of 7, $p < .10$ in the predicted direction).

Such modest results indicate that further research is needed to support the current widespread use of ratings to determine what competencies at what frequencies or levels are needed for a particular job. Also needed is research on whether ratings of individual competencies predict executive performance as well as the interview-based competency scores and algorithms do.

SUMMARY

In this article, I have reviewed research on the definition and measurement of competencies according to previously suggested principles (McClelland, 1973), and described the successful implementation, in one large multinational corporation, of a competency-assessment program based on those principles and research. The competency measures were obtained from intensive interviews detailing operant thoughts and actions associated with success in top executive occupations. The first step of this implementation was to identify competencies—carefully defined clusters of behaviors—that (at certain frequencies or levels) characterize O performers more than T performers. In the next step, these individual competencies were combined into an integrated model which suggested that most top executive positions require passing certain tipping points—that is, competency frequencies or levels that most distinguish O and T performers. Specifically, success in most of these positions required achieving the tipping point in at least 1 out of 3 individual-initiative competencies, at least 1 out of 3 organizational-skill competencies, and a total of at least 6 out of 12 competencies that either most commonly differentiate significantly between O and T managers (see Table 1) or are unique to the organization. Maximum predictive power for a competency algorithm is obtained by adding tipping-point criteria for at least 2 to 3 competencies unique to the organization involved. Sufficient research has not yet been done to determine if similar algorithms apply to success in other clusters of life outcomes, such as in health and education. But the principles of substituting competencies similar in type or function and establishing tipping points for competency levels or frequencies appear to predict success criteria across samples of people in closely aligned fields better than does the more usual method of combining predictors by multiple regression analysis.

Finally, the BEI competencies have been well enough defined for executives who get feedback on how well they meet various tipping points or competencies needed for success to improve performance in the following year.

This approach not only distinguishes employees who have been nominated as being outstanding versus typical, but also predicts who will perform better subsequently in a company (as measured by bonuses received and turnover).

Acknowledgments—This report is based on data made available by McBer & Company and by "Tastyfood," an international food and beverage company that wishes to be anonymous. The author would like to acknowledge with thanks the assistance of the staffs of these two companies.

REFERENCES

Andrews, J.D.W. (1967). The achievement motive in two types of organizations. *Journal of Personality and Social Psychology, 6,* 163–168.

Barrett, G.V., & Depinet, R.L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist, 46,* 1012–1024.

Boyatzis, R.E. (1982). *The competent manager: A model for effective performance.* New York: Wiley-Interscience.

Boyatzis, R.E., Cowen, S.S., Kolb, D.A., & Associates. (1995). *Innovation in professional education.* San Francisco: Jossey-Bass.

Caldwell, D.F., & O'Reilly, C.A., III. (1990). Measuring job-person fit with a profile comparison process. *Journal of Applied Psychology, 75,* 648–657.

Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage pregnancy. *American Journal of Sociology, 96,* 1226–1259.

Dailey, C. (1971). *Assessment of lives.* San Francisco: Jossey-Bass.

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin, 51,* 327–358.

Gladwell, M. (1996, June 3). The tipping point. *The New Yorker, 72,* 32–39.

Hogan, R., Hogan, J., & Roberts, B.W. (1996). Personality and employment decisions. *American Psychologist, 51,* 469–477.

McClelland, D.C. (1961). *The achieving society.* Princeton, NJ: Van Nostrand.

McClelland, D.C. (1973). Testing for competence rather than for "intelligence." *American Psychologist, 28,* 1–14.

McClelland, D.C. (1994). The knowledge-testing educational complex strikes back. *American Psychologist, 49,* 66–69.

McClelland, D.C., Atkinson, J.W., Clark, R.A., & Lowell, E.L. (1953). *The achievement motive.* New York: Appleton-Century-Crofts.

Motowidlo, S.J., Carter, G.W., Dunnette, M.D., Tippins, N., Werner, S., Bennett, J.R., & Vaughn, M.J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 17,* 571–587.

Nygren, D.J., & Ukeritis, M.D. (1993). *The future of religious orders in the United States: Transformation and commitment.* Westport, CT: Praeger.

Spencer, L.M., Jr., & Spencer, S.M. (1993). *Competence at work: Models for superior performance.* New York: Wiley.

Whitla, D.A. (1975). *Value added: Measuring the impact of undergraduate education.* Cambridge, MA: Harvard University, Office of Instructional Research and Evaluation.

Winter, D.G., McClelland, D.C., & Stewart, A.J. (1981). *A new case for the liberal arts: Assessing institutional goals and student development.* San Francisco: Jossey-Bass.

(RECEIVED 5/9/97; REVISION ACCEPTED 3/19/98)