CORNELL

CHR *Reports*

How to Compare Apples to Oranges

Balancing Internal Candidates' Job-performance Data with External Candidates' Selection-test Results

by Michael C. Sturman, Ph.D., Robin A. Cheramie, and Luke H. Cashen

The Center for Hospitality Research
AT CORNELL UNIVERSITY

CHR *Reports*

is produced for the benefit of the hospitality industry by

the Center for Hospitality Research at Cornell University

Cathy A. Enz, Executive Director

Glenn Withiam, Director of Publications Services

Executive Summary

How to Compare Apples to Oranges

Balancing Internal Candidates' Job-performance Data with External Candidates' Selection-test Results

by Michael C. Sturman, Ph.D., Robin A. Cheramie, and Luke H. Cashen

It has been widely accepted that past performance is a good predictor of future performance. The exact strength of that relationship, however, has been unclear. Knowing the predictive power of past performance on future performance is particularly important for employers who make hiring decisions based in part on internal candidates' performance record. Generally, some of the internal candidates' performance would be measured at different points of time (e.g., 6 months, 12 months, and 24 months ago). Others under consideration will be external candidates, whose employment information is derived from selection devices such as structured interviews and intelligence tests. This paper uses a meta-analysis to examine 20 previously published studies on the stability of job performance over time. It provides an estimate of the relationship between existing performance measures and future performance, and models the nature of this relationship as a function of the elapsed time between measures. The findings show conclusively that, in general, past performance is, indeed, a good predictor of future performance for a variety of job types (i.e., exempt, nonexempt, and those that are evaluated subjectively). Using a hypothetical selection scenario, this report also demonstrates how that information can be used to compare multiple internal and external candidates.

How to Compare Apples to Oranges: Balancing Internal Candidates' Job-performance Data with External Candidates' Selection-test Results

The growth of multiunit hospitality companies has altered the face of the lodging and food-service industries, and that change has created its share of managers' headaches regarding the need to staff chains' new properties. General shortages of qualified workers make finding motivated employees difficult.[1] Moreover, managers complain that experienced, internal workers rarely want to relocate.[2] Those needing to staff multiunit businesses are thus faced with the need to look hard both inside and outside of their companies.[3] These staffing pressures are also coming at a time when companies are reducing the number of career-advancement opportunities and are relying more on lateral transfers to fulfill their human-resources needs.[4] Taken together, those forces create a host of difficul-

[1] See: J. Choi, R.H. Woods, and S.K. Murrmann, "International Labor Markets and the Migration of Labor Forces as an Alternative Solution for Labor Shortages in the Hospitality Industry," *International Journal of Contemporary Hospitality Management,* Vol. 12, No.1 (2000), pp. 61–66. Choi *et al.* note that it is easy to find articles that report on the insufficient supply of labor for hospitality jobs, both in the United States and abroad. Much of this may be caused by general perceptions of hospitality jobs' relatively less-desirable hours and working conditions (see, for example, M. Prewitt, "A Career in Foodservice—Cons: Long Hours," *Nation's Restaurant News,* Vol. 35, No. 21 [May 21, 2001], pp. 102–103), and the perceived low pay typical of the industry (see, for example, M.C. Sturman, "The Compensation Conundrum: Does the Hospitality Industry Shortchange Its Employees—And Itself?," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 42, No. 4 [August 2001], pp. 70–76).

[2] D.J. Kennedy and M.D. Fulford, "On the Move: Management Relocation in the Hospitality Industry," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 40, No. 2 (April 1999), pp. 60–68.

[3] S. Goss-Turner, "The Role of the Multiunit Manager in Branded Hospitality Chains," *Human Resource Management Journal,* Vol. 9, No. 4 (1999), pp. 39–57. Goss-Turner's survey shows that most multiunit managers are hired from within the chain.

[4] Relocation activity within the United States is growing. Furthermore, only 34 percent of relocations are due to promotion. See: Runzheimer International, *2001 Survey & Analysis of Employee Relocation Policies and Costs* (Rochester, WI: Runzheimer International, 2001).

ties for those needing to fill staffing vacancies.

While the human-resources issues of concern to the hospitality industry are broad and complex,[5] we focus on the staffing decision that must be made when managers consider whether to hire an external candidate or make an internal transfer. This is a particularly tricky problem, because decision makers must compare one type of information on internal candidates (e.g., job-performance data) to other types of information collected on external candidates (e.g., interview results, test scores). Essentially, the person doing the selection and hiring must compare apples to oranges to make a decision.

For its use in considering external candidates, validity data abound on different selection devices.[6] However, there is no information on the validity of past-performance data for predicting future performance that would be helpful in making internal-transfer decisions. While it is practically human-resources dogma to say that past job performance is the best predictor of future job performance, it is certainly not a perfect predictor, because individual job performance changes over time. For instance, employees acquire experience, gain or lose motivation, and have different opportunities to succeed or fail. It is therefore not always clear how strongly past performance can predict future performance, let alone how this information should be viewed in comparison to information collected on external candidates (e.g., interview results, cognitive-ability tests). While companies should have information on the validity of selection devices for external candidates, the accuracy of past-performance data will likely depend on the nature of the job, how employee performance is assessed, and in particular how much time has elapsed since the most recent performance review. The manager faced with choosing between internal and external candidates can benefit from a better understanding of how to compare the "apples" of past-performance data to the "oranges" of external-selection data. In this paper, we present a model that allows one to estimate the extent that past performance predicts future performance, given **(1)** the length of time between performance measurements, **(2)** whether perfor-

[5] C.A. Enz, "What Keeps You Up at Night? Key Issues of Concern for Lodging Managers," *Cornell Hotel and Restaurant Administration Quarterly*, Vol. 42, No. 2 (April 2001), pp. 38–45.

[6] See: H.G. Heneman, R.L. Heneman, and T.A. Judge, *Staffing Organizations,* second edition (Chicago: Irwin, 1997); R.D. Gatewood and H.S. Field, *Human Resource Selection,* fifth edition (New York: Harcourt College Publishing, 2001); and F.L. Schmidt and J.E. Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," *Psychological Bulletin,* Vol. 124 (1998), pp. 262–274.

mance is measured by objective data (e.g., sales, output) or supervisory evaluations, and (3) whether the job in question is classified as exempt or nonexempt.[7] Using our model, a manager will be able to estimate the performance of transferred employees and compare them to the expected performance scores of prospective new hires.

Consider this example. The COO of multiple units of a hotel chain has to fill an assistant-GM position. The vacancy is at one of her most important locations, and thus a good hire is essential. In her search to find the best possible candidate, the COO solicits applications from both inside and outside her chain. After eliminating unqualified individuals and interviewing potential external applicants, she narrows her choice down to two finalists. The first candidate is already an assistant GM at one of the chain's other locations. This person is someone who has demonstrated above-average (but not exceptional) job performance. The other candidate is an external applicant who scored exceptionally well on the interview. So, what hiring decision should the COO make? On the one hand, past-performance data may be more valid than any kind of contrived selection device (e.g., tests, interviews). Still, how does one balance the assessment that someone *may* be an exceptional performer (i.e., the external candidate who scores well on evaluation tests) versus the known track record of an above-average, but not exceptional, performer? Moreover, what if the internal candidate's past-performance data were six weeks old? Or six months old? Or more?

We argue that in such cases hiring decisions should be made to maximize the predicted performance of the new hire.[8] To make this happen, it is essential to have an accurate estimate of the extent to which past performance predicts future performance. The goal of this paper is to provide an estimate of the ability of past performance to predict future performance, so that such ratings can be compared meaningfully against external candidates or other internal candidates. Ultimately, we show how different past-performance data can be compared

[7] The quality of a job being "exempt" or "nonexempt" refers to whether employees in the job are covered by the Fair Labor Standards Act. More information about classifying jobs is available from the Wage and Hour Division, Employment Standard Administration, U.S. Department of Labor. (See: www.dol.gov/esa/whd/.)

[8] We recognize that there may be other goals when making hiring decisions. For example, an internal transfer may be made for developmental reasons; or, individuals that promise a better fit with an organization's culture may be hired over otherwise more-qualified candidates. Nevertheless, obtaining top performers is a major concern, and we worked under the assumption that that is the fundamental goal when making hiring decisions.

against data on external candidates to help make hiring decisions.

The Accuracy of Past-performance Data

We should expect past performance to predict future performance, but there are a number of reasons to expect that such is not always necessarily the case. The influences of these changes can be categorized as either **(1)** changes in the individual or **(2)** changes in the job.

Why performance changes over time. Individuals change over time. Most obviously, employees gain job experience that generally helps them to perform better, know how to "get things done" within the company, and know how to do their job more efficiently. Individuals may also change their levels of motivation over time. Some employees may become disillusioned with a particular job, or even the company. They may not work as hard, may be looking for other employment opportunities, or simply may not care about their performance. Others may become more motivated—for instance, incentive plans may make some employees work harder or a new manager may be inspirational or otherwise lead employees to better performance. Training opportunities may give employees new skills to succeed and motivate employees to use those skills to enhance their performance levels.

Changes in the job may also affect the stability of individual performance levels. In the hospitality industry, jobs must change to reflect the changing demands of customers. Changes in the organizational climate, local conditions, and national economy may affect the expectations of individual job performance, and therefore evaluations of competence. For example, an across-the-board decline in occupancy may mean that hotel managers will ask their employees to focus on delivering excellent customer service, as the need for efficiency becomes less critical. However, if economic conditions change such that demand rises sharply and a particular hotel has not increased staffing levels to accommodate that demand, the manager may nevertheless consider worker efficiency to be of utmost importance when evaluating individual job performance. Jobs may also change through the introduction of new services, the use of new technologies, or the redesign or combination of tasks. Even if employee characteristics were to remain stable over time, a changing job may lead to changes in individuals' performance.[9]

---

[9] For example, see: J. Perdue, R. Woods, and J. Ninemeier, "Competencies Required for Future Club Managers' Success," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 42, No. 1 (February 2001), pp. 60–65; and S. Formica and K. McCleary, "Professional-development Needs in Italy," *Cornell Hotel and Restaurant Administration Quarterly,* Vol. 41, No. 2 (April 2000), pp. 72–78.

Similarly, even if the nature of the job itself remains the same, the measurement of what constitutes performance on the job may change. For example, a firm's performance-appraisal system may be redesigned (or a formal performance-appraisal system may be introduced for the first time). If performance is assessed in new ways, or if the relative importance of different components of job performance changes, then individual performance is also likely to change. In part, performance may simply appear more dynamic as the measurement system changes. However, if employees are aware of the new measurement system, it is likely that they will devote their attention to the more important (or at least, the most heavily rewarded) components of the job.

Of course, it is reasonable to expect that individuals, jobs, and measurement systems all change over time. Jobs may require the use of new technologies, individuals may receive training on that technology, and appraisal systems are constantly updated. Some people will be more adept at the new tools, or have more relevant experience from prior jobs. Similarly, some people will learn more from a particular training course than will others. New managers may come onto the scene and have different views of the appraisal system or different priorities than did previous managers. Thus, hospitality jobs exist in a dynamic environment—and as people learn, grow, and change, it should not be surprising that evaluations of individual performance will also change over time.

It is important to note that as time passes, more change may occur. Individual ability and motivation, measured from week to week or even month to month, are unlikely to change dramatically. Similarly, from one day to the next, jobs have relatively stable overall duties and requirements. Generally, on a day-to-day basis, there is not enough time to accumulate substantial experience to influence one's knowledge, skills, or abilities, or to cause a fundamental shift in the nature of one's work. However, over the course of a year, and especially many years, individuals and jobs may change dramatically. All the individual and job changes described above are more likely to occur over time frames measured by years versus days or weeks.

Reasons for Performance Consistency

Changes in the individual and the job make it clear that individual performance will change over time. However, there are characteristics of individuals and jobs that do remain reasonably stable, and therefore might suggest a level of consistency in individual job performance no matter how long the time between measures of performance.

One such constant influence is cognitive ability. An individual's cognitive ability has been shown to be a highly valid predictor of individual job performance, and cognitive ability is important for individuals when adapting to changes in jobs' demands.[10] Moreover, psychological research has shown that cognitive ability remains relatively stable over individuals' working lives. We would thus expect that some aspects of individual job performance will remain constant. Indeed, Farrell and McDaniel showed that cognitive ability plays a role in predicting individual performance both in the short and the long term (even after 10 years on a job).[11]

Another stabilizing influence on individual job-performance scores over time is personality. Personality plays a role in individuals' approaches to employment, their work ethic, and their general dispositions at the workplace. Like cognitive ability, individuals' personality also remains relatively stable over adults' lives and is related to individual job performance.[12] Thus, some portion of individual job performance that is directly relevant to personality should remain relatively constant over people's working lives.

---

[10] Murphy proposed that a person's tenure in an organization includes two distinct stages: transition stage and maintenance stage. During the transition stage the job is new to the employee, who must rely on cognitive ability to learn new tasks and solve new problems. Once workers learn the job, they enter the maintenance stage. During this stage, task performance is attributable to the performance of well-learned processes, and thus cognitive ability plays a minor role in individual performance. However, as jobs almost always include some need to adapt to changes, cognitive ability should always play some role in the prediction of individual job performance. See: Kevin Murphy, "Is the Relationship between Cognitive Ability and Job Performance Stable Over Time?," *Human Performance,* Vol. 2 (1989), pp. 183–200; P.L. Ackerman, "Determinants of Individual Differences during Skill Acquisition: Cognitive Abilities and Information Processing," *Journal of Experimental Psychology: General,* Vol. 177 (1988), pp. 288–318; and J.N. Farrell and M.A. McDaniel, "The Stability of Validity Coefficients Over Time: Ackerman's (1988) Model and the General Aptitude Test Battery," *Journal of Applied Psychology,* Vol. 86 (2001), pp. 60–79.

[11] Farrell and M.A. McDaniel, *op. cit.*

[12] See: T.A. Judge, C.A. Higgins, C.J. Thoresen, and M.R. Barrick, "The Big Five Personality Traits, General Mental Ability, and Career Success Across The Life Span," *Personnel Psychology,* Vol. 52 (1999), pp. 621–652. Judge *et al.* show that childhood ratings of personality characteristics were related to measures of career success measured in late adulthood. As personality characteristics are expected to maintain their relationships with job performance as they do over time with measures of career success, a portion of individual job performance should consistently be able to be predicted by one's personality.

Consistency of Performance: What to Expect

The above discussion suggests that performance changes over time, but some characteristics of individuals may continue to predict future performance no matter how much time passes. Given that individuals often have their job performance evaluated at different times (e.g., semiannually, annually, biennially), and performance evaluations can be skipped (or lost), it may be necessary to be able to compare past-performance levels of individuals when notably different amounts of time have elapsed since the last performance assessment. Furthermore, it may be necessary to compare past-performance data to information collected from new applicants (such as that obtained from a structured interview, mental-ability test, or work simulation).

The first step in helping managers compare past-performance data with external candidates' data is to detail the relationship between an employee's measures of performance taken at different times.

The goal of these analyses will be to provide specific evidence of the accuracy of past job-performance data for predicting future performance, so that existing evidence of an employee's job performance can be used to make strategic selection decisions—especially to compare individuals' different types of performance information (e.g., performance data from another time period, external-applicant information).

The Stability of Job-performance Ratings

Recall that our goal is to provide an estimate of the correlation between past performance and future performance in the same job. When estimating the accuracy of traditional selection devices—such as an interview or cognitive-ability test—a single correlation coefficient is estimated to represent the strength of the relationship between the selection devices' scores and job performance ratings. However, as already discussed, there does not exist one single correlation representing the strength of the relationship between individuals' job performance ratings at different times. Rather, the relationship will in part depend on how much time has passed between performance measurements. We thus need our analyses to model the strength of the relationship between performance ratings taken at different points of time.

To assess the stability of job-performance ratings over time, we relied on the statistical technique of meta-analysis. Meta-analysis allows one to empirically combine previous studies examining a particular relationship. It is useful because it allows an investigator to combine vast amounts of previously published data, yielding a sample larger than that used in any single study.

One complication of measuring the stability of job performance over time is that it is difficult to separate random changes in performance from systemic changes in performance. For example, if performance between two months is correlated at 0.80, it is not clear whether the lack of perfect stability is due to flaws in the evaluation process or to actual changes in the individual's performance. In reality, performance changes are probably due to both phenomena. Therefore, when examining job-performance levels over time, it is necessary to use previous studies that have measured performance in more than two time periods.[13]

Previous research. We conducted a four-stage search for articles that studied individual performance over three or more time periods. First, we used articles that reviewed literature on dynamic performance as a source of potential studies.[14] Second, we performed a manual search of the following management and marketing journals: *Journal of Applied Psychology, Academy of Management Journal, Administrative Science Quarterly, Personnel Psychology, Organizational Behavior and Human Decision Processes, Journal of Management, Human Resource Management, Human Relations, Journal of Marketing,* and *Journal of the Academy of Marketing Science.* Third, we conducted a computer search using *ABI Inform*, which contains abstracts and articles regarding business and psychological research. Fourth, we solicited unpublished manuscripts (and manuscripts not revealed through our prior search) through e-mail listserves associated with the Academy of Management.[15]

---

[13] With data from multiple time periods, we assume that the amount of unreliability (or noise) at each time period is roughly equal. Thus, when we have at least three observed relationships between performance measures, we can mathematically estimate the extent to which a lack of perfect stability is attributable to **(a)** the noise in each period and **(b)** other changes in the constructs being measured. Three performance measurements are required to make this estimate; however, more accurate estimates can be made as the number of performance measurement periods increases.

[14] See: G.V. Barrett, M.S. Caldwell, and R.A. Alexander, "The Concept of Dynamic Criteria: A Critical Reanalysis," *Personnel Psychology,* Vol. 38 (1985), pp. 41–56; C.L. Hulin, R.A. Henry, and S.L. Noon, "Adding a Dimension: Time as a Factor in the Generalizability of Predictive Relationships," *Psychological Bulletin,* Vol. 107 (1990), pp. 328–340; and W.W. Rambo, A.M. Chomiak, and J.M. Price, "Consistency of Performance Under Stable Conditions of Work," *Journal of Applied Psychology,* Vol. 68 (1983), pp. 78–87.

[15] "Human Resource Division Net" and "Research Methods Net," both listserves associated with their respective divisions of the Academy of Management (see: www.aom.pace.edu).

EXHIBIT 1

Individual performance studies included in the meta-analysis

Cite	Number of correlations	Mean (N)	Min. time	Max. time	Min. *r*	Max. *r*	Mean *r*	Sample
Adkins and Naumann (2001)	15	214	1	5	0.34	0.67	0.53	Telephone sales agents
Bass (1962)	6	99	12	42	0.29	0.58	0.43	Food sales people
Breaugh (1981)	6	101	12	36	0.50	0.69	0.5	Research scientists
Deadrick and Madigan (1990)	15	413	1	5	0.66	0.92	0.82	Sewing-machine operators
Griffin (1991)	6	545	6	48	0.38	0.65	0.52	Bank tellers
Hanges, Schneider, and Niles (1998)	78	79	6	72	0.05	0.62	0.35	Faculty members (teaching)
Harris, Gilbreath, and Sunday (1998)	3	218	12	24	0.24	0.56	0.40	Government: contract workers, and clerical, technical, managerial, and professional jobs
Harrison, Virick, and William (1996)	11	154	1	11	-0.13	0.55	0.34	Sales representatives
Hoffman, Nathan, and Holden (1991)	3	62	12	24	0.74	0.84	0.78	Service jobs in a utility company
Hofmann, Jacobs, and Baratta (1993)	66	319	3	33	-0.05	0.63	0.27	Insurance sales personnel
McEvoy and Beatty (1989)	3	64	12	24	0.61	0.71	0.65	Managers
Mitchel (1975)	3	128	36	60	0.77	0.93	0.83	Managers
Ployhart and Hakel (1998)	28	303	3	21	0.29	0.68	0.46	Securities brokers
Ravlin, Adkins, and Meglino (1994)	3	167	12	36	0.29	0.46	0.37	Production workers
Reilly, Smither, and Vasilopoulos (1996)	6	92	6	30	0.32	0.57	0.46	Managers
Rothe (1947)	3	130	0.5	1	0.57	0.72	0.66	Machine workers
Rothe (1970)	11	22	0.25	2.75	0.14	0.73	0.45	Welders
Russell (2001)	3	98	12	36	0.55	0.66	0.60	General managers
Sturman and Trevor (2001)	28	724	1	7	0.38	0.55	0.48	Loan originators
Warr and Bunce (1995)	3	106	3	7	0.65	0.72	0.67	Junior managers

Note: The number of correlations represents the number of performance–performance correlations available in each study. The Mean N is the average sample size used in each study. Time is measured in months, with “Min. time” representing the smallest elapsed time in the study when measuring the performance–performance relationship, and “Max. time” representing the longest elapsed time. “Min. *r*,” “Max. *r*,” and “Mean *r*” report the range of and average (uncorrected) correlations from each study.

From those sources we found 20 papers that examined individual performance over three or more time periods; those studies were included in our meta-analysis. A summary of the sample and characteristics of each of those 20 studies is provided in Exhibit 1, as is the range of correlation coefficients and the average correlation reported in each study.

Before performing any analyses, we corrected each correlation for range restriction and unreliability.[16] Because we expect that the length of time between measures will influence the recorded job performance, we modeled the extent to which performance was related to time (measured in months). To consider other factors that might influence performance over time, we also controlled for whether the performance evaluation was a subjective supervisory rating or an objective measure of output (e.g., sales), and whether the job in question was classified as exempt or nonexempt (because the measurement criteria for exempt jobs are different from those of nonexempt jobs).

We used multiple regression to estimate the model approximating the correlation of performance scores across different time lags. Although there are many ways to perform a meta-analysis, this technique is relatively easy to understand compared to other techniques, and

[16] Range restriction and unreliability are both characteristics of the performance-measurement process that may limit the accuracy of the performance assessment, and thus also decrease the apparent relationship between items measured. A full discussion of the mathematical and practical implications of measures with range restriction and unreliability can be found in: J.E. Hunter and F. Schmidt, *Method of Meta-Analysis: Correcting Error and Bias in Research Findings* (Newbury Park, CA: Sage, 1990). In short, range restriction occurs when people at the top or bottom of the distribution are eliminated. For example, if performance is rated on a 5-point scale, but all bad performers are fired so that only those with scores of 4 or 5 remain, then observed relationships between performance and other characteristics will appear much weaker than would the true relationship. When considering performance over time, low performers tend to leave organizations (or are dismissed), thus restricting the range of observed performance scores and making the relationship between performance measures over time appear artificially unstable. Fortunately, a number of researchers have determined ways of "correcting" observed correlations for the problems of range restriction. For a detailed account, see: P.R. Sackett and H. Yang, "Correction for Range Restriction: An Expanded Typology," *Journal of Applied Psychology,* Vol. 85 (2000), pp. 112–118. Unreliability is the amount of random error that occurs in measurement, such as from the subject being distracted, being unable to observe all relevant actions, and being rushed. Again, scholars have developed ways to measure the extent of unreliability and correct observed correlations for this phenomenon (for example, see: Hunter and Schmidt, *op. cit.*). By correcting for range restriction and unreliability, we can calculate a more accurate estimate of the relationship between variables over time. This estimate will then be comparable to estimates of the validity of other selection devices, such as a structured interview and a cognitive-ability test.

EXHIBIT 2

Results of the meta-analyses

	Comparison		
	I	II	II
	No covariates	With time covariate	With time covariate and other controls
$ß_0$ [Intercept]	0.508 (0.0116)*	0.715 (0.0334)*	0.948 (0.0295)*
$ß_1$ [Supervisory ratings = 1; otherwise 0]	—	0.015 (0.0272)	0.220 (0.0246)*
$ß_2$ [Exempt = 1; nonexempt = 0]	—	-0.2384 (0.0339)*	-0.180 (0.0258)*
$ß_3$ [(ln (time+1))]	—	—	-0.159 (0.0105)*
% Total Variance Explained	—	16%	52%

Notes: * $p < .0001$. Analyses based on 20 samples, 300 correlations, and a total of 75,708 observations of individual job performance. Each comparison model is significantly more predictive than the previous comparison model (at $p < .0001$). Note that beta-coefficients are used to predict the arc-tangent of the correlation. To compute the estimated correlation, the estimated value needs to be transformed using a hyperbolic tangent function (commonly represented as TanH).

it yielded essentially the same answer as more-complex techniques.[17] In essence, we predicted the correlation between performance measures over time as a function of **(1)** the type of performance measure, **(2)** the type of job, and **(3)** the length of time between performance measures.[18]

Analyzing Job-performance Ratings

Exhibit 2 reports the results of the regression analyses, and Exhibit 3 shows the estimated relationship between performance scores over time. The regression model predicted over 50 percent of the variance in the correlation coefficient, and thus represents a highly predictive model explaining the range of correlations found between performance measures.[19]

Note that the graph in Exhibit 3 shows the modeled relationship between performance scores for various time lags for **(1)** exempt jobs with subjective performance evaluations, **(2)** exempt jobs with objective performance evaluations, **(3)** nonexempt jobs with subjective perfor-

[17] We also performed the analysis using a technique called hierarchical linear modeling. The derivation of this meta-analytic method is provided in detail in: M.C. Sturman, R. Cheramie, and L. Cashen, “Consistency, Stability, and Test-retest Reliability of Employee Job Performance: A Meta-analytic Review of Longitudinal Findings,” Center for Hospitality Research working paper, 04-08-01. Although this technique has a number of advantages from the point of view of statistical correctness, it yielded essentially the same results as the simpler technique reported here. Thus, we chose to report the simpler technique to facilitate the interpretation of our results.

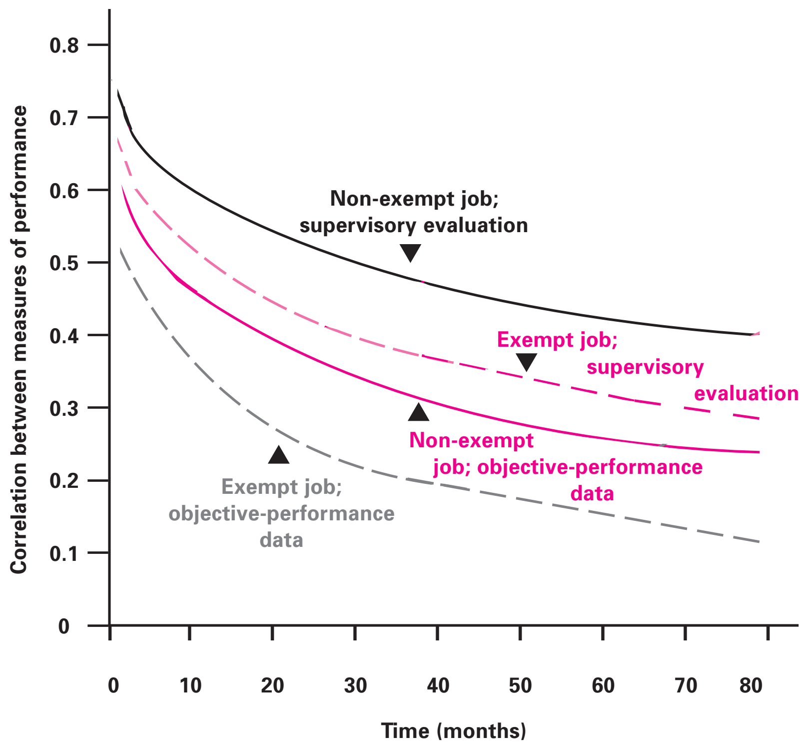
[18] The type of performance measure was coded as 1 for supervisory ratings, and 0 for measures such as output or sales. For the type of job, exempt jobs were coded as 1, and nonexempt jobs were coded as 0. Time was measured in months, but was transformed with a natural logarithm to better meet the statistical assumptions of regression analysis.

[19] We also performed statistical tests to determine whether the amount of error that we observed was more than we would have expected by chance. This test-of-homogeneity test was not significant.

mance evaluations, and (4) nonexempt jobs with objective performance evaluations. Based on a variety of alternative models measuring time in different ways (the natural logarithm of the number of months between performance measures plus one, time as a linear variable, and time and time-squared as variables), we found in all cases that (1) as the elapsed time between performance ratings increased, the relationship between those performance ratings decreased, (2) the relationship between performance ratings did not reach zero, but rather a value greater than zero, (3) the relationship between performance ratings depended on the job type and the way performance was measured, and (4) the apparent stability of performance scores at a time of zero was a value less than 1.0. The fourth result suggests that there exists some test-retest unreliability in the measurement of job performance. We can approximate the test-retest reliability of performance scores over time by estimating the correlation between performance scores with a hypothetical time lag of zero months.[20] We will

EXHIBIT 3

Relationship between performance measures over time (observed and corrected for test–retest unreliability)



[20] The level of unreliability was estimated as the estimated consistency of job performance with a hypothetical time lag of zero months. The estimated "true" stability could thus be approximated by taking the predicted value at any given time lag and dividing this value by the estimated reliability.

- For an exempt job with supervisory ratings, the reliability = TanH(0.988) or 0.757.
- For an exempt job with objective ratings, the reliability = TanH(0.768) or 0.646.
- For a nonexempt job with supervisory ratings, the reliability = TanH(1.168), or 0.824.
- For a nonexempt job with objective ratings, the reliability = TanH(0.948), or 0.739.

EXHIBIT 4

Hypothetical assistant-general-manager-applicant characteristics

Applicant	Type of candidate	Type of information on candidate	Age of performance data	Score
Applicant 1	Internal	Past performance	6 months	5.0
Applicant 2	Internal	Past performance	6 months	4.5
Applicant 3	Internal	Past performance	6 months	4.0
Applicant 4	Internal	Past performance	12 months	5.0
Applicant 5	Internal	Past performance	12 months	4.5
Applicant 6	Internal	Past performance	12 months	4.0
Applicant 7	Internal	Past performance	24 months	5.0
Applicant 8	Internal	Past performance	24 months	4.5
Applicant 9	Internal	Past performance	24 months	4.0
Applicant 10	External	Structured interview	N.R.	5.0
Applicant 11	External	Structured interview	N.R.	4.5
Applicant 12	External	Structured interview	N.R.	4.0

Past-performance and structured-interview scores are all calculated using a Likert-type 1-to-5 scale, with 5 indicating the best possible score. The average of all scores is assumed to equal 3, with a standard deviation of 1. The above list represents the final set of job candidates in this simplified hiring example.

use this value below to provide a more accurate estimate of the true stability of performance scores over various time lags.

The Implications of the Statistics

Our efforts provide a model of the predictive value of past-performance data on future performance (given the same job). Not surprisingly, we also learned that the predictive value of performance measures becomes weaker as time passes. However, by having a specific estimate of the extent of that erosion, past-performance data can be compared to other selection information, thus allowing for more-intelligent staffing decisions.

Recall the example of a multiunit hotel-chain franchisee who is opening a new facility and is looking for a new assistant general manager. She is considering a number of incumbent employees from her other properties as well as several external applicants. The incumbent employees can all present past-performance data, although from different time periods; the external applicants have each undergone a structured interview. Exhibit 4 shows a set of hypothetical applicants for this scenario, and the type of information collected on each. In this example, performance was rated using a Likert-type scale of 1 to 5, with 5 being high.[21] As Exhibit 4 shows, there are candidates ranging from high performers (5.0) to above-average performers (4.0). For the

[21] For illustrative purposes, we are also saying that individuals' interview scores are rated on a 1-to-5 scale, with 5 being high. For both performance ratings and interview score, we are assuming that the average (mean) score is a 3, and the standard deviation is 1. Furthermore, because this is a hypothetical hiring decision, we have already narrowed the search down to the top candidates (those with scores of 4 or higher). The remaining question is, how to rank the remaining candidates. We will also assume that the structured interview has a validity of 0.40.

internal applicants, the lengths of time since the last performance assessment were 6 months, 12 months, and 24 months. Given only the simplified information provided for this example, how should the franchisee rank the expected future performance levels of all 12 candidates?

There are some obvious answers. Clearly, of the external candidates (10, 11, and 12), the one scoring a 5.0 on the structured interview is better than the one scoring a 4.5, and both are better than the one scoring a 4.0. But how do the external candidates compare to the internal candidates? Predicted performance scores can be computed by multiplying the standardized score on the selection device (either the past-performance data or the structured interview) by the accuracy of the selection device.[22]

For the external candidates, the validity will be assumed to be 0.40;[23] for the internal candidates, the strength of the relationship must be determined based on the results reported in Exhibit 2.

To use the results from Exhibit 2 for estimating the accuracy of existing performance ratings in predicting future performance, you need three pieces of information: **(1)** the length of time between performance measures (measured in months), **(2)** whether performance is rated using subjective evaluations or objective performance measures, and **(3)** whether the job is exempt or nonexempt. With those data, the correlation is estimated using one of the four formulas shown at the top of the next page.[24]

---

[22] For example, for Applicant 1, the individual's past-performance score was a 5. This value is converted to standardized units by subtracting the mean (3) and dividing the result by the standard deviation (1). Thus, Applicant 1's standardized performance score is 2.0. This value is then multiplied by the validity of the selection information (as shown in Exhibit 5, it is 0.76), suggesting a predicted job performance of 1.52. This number is then converted back into the 1-to-5 scale by multiplying it by the standard deviation (1) and adding the mean (3). The resulting predicted performance score of Applicant 1 is thus 4.52. This same method is applied for each of the job applicants, based on their current information (shown in Exhibit 2) and the validity of that information (shown in Exhibit 3).

[23] Evidence suggests that the validity of the structured interview is around 0.40. McDaniel *et al.*, using a meta-analysis of 89 studies, estimated the true validity to be 0.44. See: M.A. McDaniel, D. Whetzel, F. Schmidt, and S. Maurer, "The Validity of Employment Interviews: A Comprehensive Review and Meta-analysis," *Journal of Applied Psychology,* Vol. 79 (1994), pp. 599–616. A meta-analysis by Schmidt and Rader, based on 33 different studies, estimated the validity of the structured interview to be 0.40. See: F.L. Schmidt and M. Rader, "Exploring the Boundary Conditions for Interview Validity: Meta-analytic Validity Findings for a New Interview Type," *Personnel Psychology,* Vol. 52 (1999), pp. 445–464.

[24] Note that the denominator in each formula is the estimated reliability, reported earlier from footnote 21.

EXHIBIT 5

Correlation between past-performance and future-performance scores for four job types and for various elapsed-time periods

Elapsed time	Nonexempt; objective ratings	Nonexempt; subjective ratings	Exempt; objective ratings	Exempt; subjective ratings
1 month	0.93	0.89	0.95	0.93
2 months	0.89	0.82	0.92	0.88
3 months	0.85	0.77	0.90	0.84
4 months	0.83	0.73	0.88	0.81
5 months	0.80	0.69	0.86	0.79
6 months	0.78	0.66	0.84	0.76
9 months	0.73	0.59	0.81	0.71
12 months	0.69	0.53	0.78	0.67
18 months	0.63	0.45	0.73	0.60
24 months	0.59	0.39	0.70	0.56
30 months	0.55	0.34	0.67	0.52
36 months	0.52	0.30	0.65	0.48
48 months	0.47	0.23	0.61	0.43
60 months	0.43	0.18	0.57	0.39
72 months	0.39	0.13	0.55	0.35

- For an exempt job with subjective ratings, the correlation =

  TanH {0.988 – 0.159 × [ln (time + 1)]} ÷ 0.757

- For an exempt job with objective ratings, the correlation =

  TanH {0.768 – 0.159 × [ln (time + 1)]} ÷ 0.646.

- For a nonexempt job with subjective ratings, the correlation =

  TanH {1.168 – 0.159 × [ln (time + 1)]} ÷ 0.824.

- For a nonexempt job with objective ratings, the correlation =

  TanH {0.948 – 0.159 × [ln (time + 1)]} ÷ 0.739.

Exhibit 5 shows the estimated relationship between performance scores for the four types of jobs for a number of different time lags.

Exhibit 6 shows the correlations representing the predictive accuracy of the selection devices and the predicted performance scores of the 12 candidates, along with the ranking of the candidates that results from those calculations. Our results show that the strong relationship between past performance and future performance should encourage employers to rely most heavily on this information. Indeed, it can be seen that the top candidate based on the interview score is ranked only seventh overall among all 12 candidates. Likewise, the individual who scored 4.5 (well above average) on the interview is nevertheless ranked lower than all but one of the internal candidates. This occurs because past performance is such a good predictor of future job performance in exempt jobs that receive subjective ratings. Because structured interviews are not nearly as good at predicting future performance as is actual performance data, an employer is better off taking the above-average candidate who has more known information than the candidate who appears exceptional on the (less valid) structured interview.

Note, however, that not all performance measures are equally valid when predicting future performance for different types of jobs. Exhibit 7 (on the next page) shows candidates for another hypothetical opening: a sales position. For exempt jobs where performance is easily quantified (e.g., the amount of sales), the accuracy of past-performance data is not as high as in the assistant-GM example. For the sales-position scenario, we considered standardized performance scores of 2, 1.5, and 1 standard deviation above the mean. Using the same techniques described in the previous example, but using the correlation values appropriate for a sales position, it is apparent that the selection decision would be somewhat different than that of the assistant-GM example. As shown in

EXHIBIT 6

Predicted performance scores of hypothetical applicants for assistant-general-manager position

Applicant	Score	Correlation between selection device and job performance	Predicted performance	Rank
Applicant 1	5.0	0.78	4.52	1
Applicant 2	4.5	0.78	4.17	4
Applicant 3	4.0	0.78	3.78	8
Applicant 4	5.0	0.69	4.38	2
Applicant 5	4.5	0.69	4.04	5
Applicant 6	4.0	0.69	3.69	9
Applicant 7	5.0	0.59	4.18	3
Applicant 8	4.5	0.59	3.89	6
Applicant 9	4.0	0.59	3.59	11
Applicant 10	5.0	0.40	3.80	7
Applicant 11	4.5	0.40	3.60	10
Applicant 12	4.0	0.40	3.40	12

For the internal applicants (1–9), we are assuming that past performance was measured using subjective ratings. Assistant-GM positions are also classified as exempt. Validity of the structured interview is 0.40; the relationship between past-performance ratings and future-performance ratings is based on the results shown in Exhibit 2. A score of 4.0 (on a scale of from 1 to 5, with a mean of 3 and standard deviation of 1) is 1 standard deviation above the mean. A score of 4.5 is 1.5 standard deviations above the mean, and a score of 5.0 is 2 standard deviations above the mean.

EXHIBIT 7

Hypothetical sales-applicant characteristics

Applicant	Type of candidate	Type of information on each candidate	Age of performance data	Number of standard deviations above mean
Applicant 1	Internal	Past sales	6 months	2
Applicant 2	Internal	Past sales	6 months	1.5
Applicant 3	Internal	Past sales	6 months	1
Applicant 4	Internal	Past sales	12 months	2
Applicant 5	Internal	Past sales	12 months	1.5
Applicant 6	Internal	Past sales	12 months	1
Applicant 7	Internal	Past sales	24 months	2
Applicant 8	Internal	Past sales	24 months	1.5
Applicant 9	Internal	Past sales	24 months	1
Applicant 10	External	Structured interview	N.R.	2
Applicant 11	External	Structured interview	N.R.	1.5
Applicant 12	External	Structured interview	N.R.	1

Past-performance and structured-interview scores are all represented by standardized units (i.e., a scale with a mean of 0 and a standard deviation of 1).

Exhibit 8, because performance is not as good a predictor of future performance in this case, high structured-interview scores are more likely to indicate accurately whom to hire. Although the external candidates are still not the most desirable hires for the sales position, the top two external candidates now are ranked fourth and eighth, respectively, among all 12 applicants, and the lowest rated of the 12 candidates is an internal applicant. While some of the results in this second example are similar to those found in the assistant-GM example, the specific differences could ultimately lead to different hiring decisions.

Suggestions for Management

Our study confirms a number of opinions that have long survived as human-resources principles, namely, that performance changes over time, and that past performance is the best predictor of future performance. Moreover, what's new from this study is **(1)** the specific documentation of the extent to which those two statements are true, and **(2)** the specific techniques that can be used to compare internal and external candidates.

To help prepare managers to make better hiring decisions when comparing internal and external candidates, the following steps should be taken.

(1) Collect information. It is essential to gather individual-performance data regularly, and keep all the data in a human-resources information system. Make sure that this information is collected in the same manner from each of your units, and is entered into a global database. The data will do little good if they cannot be used to compare individuals across units.

(2) Estimate the accuracy of past-performance data. Look at the relationship between performance ratings taken at different points of time, particularly for those jobs that might involve lateral transfers. If your human-resources professionals have the tools, ability, and data, they can use records of past performance to estimate the accuracy of past performance to predict future performance. Otherwise, the estimates from this study can be employed (using the formulas above, or the results in Exhibit 5).

(3) Collect data on external candidates. When collecting data on external candidates, use a valid selection system, and document the accuracy of the system. Perform a validity study if possible; otherwise, rely on published research or validity studies performed by the developer of the selection device.

(4) Make fair comparisons. Use the statistical evidence from above, in conjunction with individuals' scores, to make a fair comparison

EXHIBIT 8

Predicted performance scores of hypothetical applicants for sales positions

Applicant	Score (SDs)	Correlation between selection device and job performance	Predicted performance	Rank
Applicant 1	2.0	0.66	1.32	1
Applicant 2	1.5	0.66	0.99	3
Applicant 3	1.0	0.66	0.66	7
Applicant 4	2.0	0.53	1.06	2
Applicant 5	1.5	0.53	0.795	5
Applicant 6	1.0	0.53	0.53	10
Applicant 7	2.0	0.39	0.78	6
Applicant 8	1.5	0.39	0.59	9
Applicant 9	1.0	0.39	0.39	12
Applicant 10	2.0	0.40	0.80	4
Applicant 11	1.5	0.40	0.60	8
Applicant 12	1.0	0.40	0.40	11

For the internal applicants, we are assuming that past performance was measured using sales figures (hence, an objective rating). We also assume that the sales positions are classified as exempt. Validity of the structured interview is 0.40; the relationship between past-performance and future-performance ratings is based on the results shown in Exhibit 2. Scores are already expressed in standardized units.

across candidates. Rely on the statistical evidence, and not "gut" impressions that contradict the estimates of actual validity. It may be tempting to hire the candidate who performs superbly in the interview over the "only" above-average candidate as indicated by past-performance data. Relying on the statistical evidence will yield the greatest probability of maximizing employee job performance.

If the goal of the job-selection process is to hire those who will be the best performer, then hiring decisions should be made based on whatever tool provides the most accurate prediction of future performance. To do that fairly, the information on candidates should be weighted by the accuracy of that information. Past performance is a good predictor of future performance, and valid performance data should be used to make hiring decisions. Moreover, that prediction holds over long time lags. The practical significance of that relationship appears even stronger when it is compared to other highly recommended selection devices, such as the structured interview, cognitive-ability test, and job simulations.[25]

This means that, in many cases, above-average candidates for whom there is reliable, valid information should be selected over those top candidates who perform exceptionally well on less-valid instruments.

Limitations. Of course, the results of our study are limited by the fact that our analyses are based on a broad set of jobs and samples (so that we could generalize the findings). While using such a broad brush is advantageous in that it provides strong evidence that the relationship between past and future performance is similar across a number of different jobs and employment circumstances, it may be fruitful for companies to perform their own analyses to further refine our results. Specifically, companies that collect and keep performance records can determine exactly the stability of workers' performance scores over time and for specific jobs. The collection of validity data to support hiring and promotion decisions is essential for developing a modern strategic human-resources program.[26] ❑

---

[25] Heneman *et al.* state that validities above 0.30 are of high usefulness. Moreover, as shown in Schmidt and Hunter's article (*op. cit.*), even the best validities for selection devices are rarely above 0.50.

[26] See: Heneman *et al., op. cit.* and Gatewood and Field, *op. cit.*

About the Researchers

Michael C. Sturman, Ph.D. (top photo), is an associate professor at the Cornell University School of Hotel Administration (mcs5@cornell.edu) and editor of *Cornell Hotel and Restaurant Administration Quarterly.*
Robin A. Cheramie (middle photo; rchera2@lsu.edu) and
Luke H. Cashen (bottom photo; lcashen@lsu.edu) are doctoral students in Louisiana State University's department of management.

About The
Center for Hospitality Research
at Cornell University

The Center for Hospitality Research is the premier source for in-depth research on the hospitality industry. The Center constitutes a community of researchers and industry sponsors working in concert to develop and dispense new knowledge that directly bears on the hospitality industry. Membership in the Center is open to any individual or corporation interested in furthering high-quality research for the betterment of the hospitality industry.

Our corporate sponsors include the following firms.

Bartech Automatic Systems

Cornell Hotel Society Foundation

Four Seasons Hotels and Resorts

Richmond Events

Willowbend Golf Management

Windsor Hotels

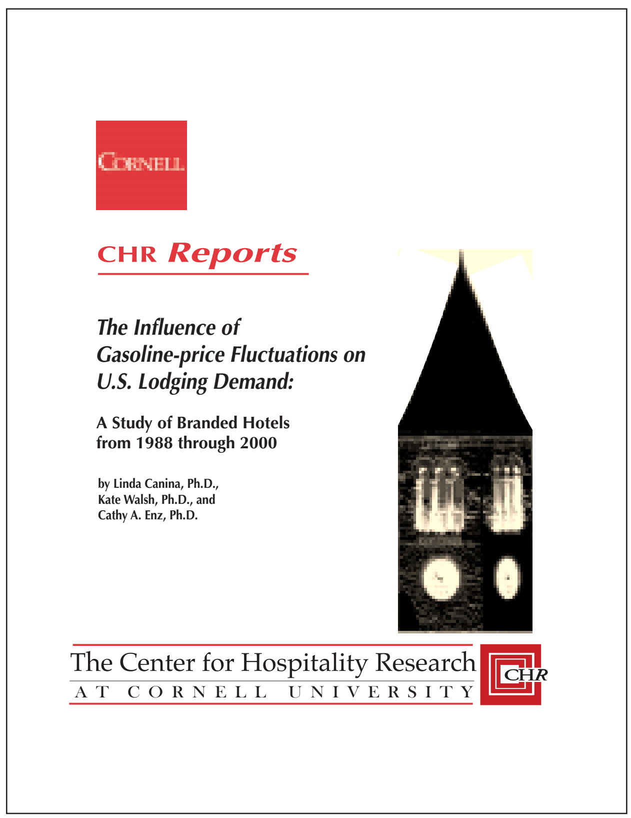
Also available from The Center for Hospitality Research

Strengthening the Purchaser–Supplier Partnership: Factors that Make a Difference

Cornell professors Judi Brownell and Dennis Reynolds examined the factors that cemented strategic partnerships among food-service operators and their suppliers. Key factors include trust, service, and communication skills.

The Influence of Gasoline-price Fluctuations on U.S. Lodging Demand

This study quantifies the effects on room demand of changes in gasoline prices and explores the ramifications of those changes.



For information on these and other CHR publications, visit our web site (www.hotelschool.cornell.edu/chr), or telephone 607-255-9780.

The Center for Hospitality Research

AT CORNELL UNIVERSITY