B

BALANCED SCORECARD

Balanced scorecard is a management system that enables organizations to translate vision and strategy into action. This system provides feedback on internal business processes and external outcomes to continually improve organizational performance and results. Robert Kaplan and David Norton created the balanced scorecard approach in the early 1990s.

Most traditional management systems focus on the financial performance of an organization. According to those who support the balanced scorecard, the financial approach is unbalanced and has major limitations:

- Financial data typically reflect an organization's past performance. Therefore, they may not accurately represent the current state of the organization or what is likely to happen to the organization in the future.
- It is not uncommon for the current market value of an organization to exceed the market value of its assets. There are financial ratios that reflect the value of a company's assets relative to its market value. The difference between the market value of an organization and the current market value of the organization's assets is often referred to as intangible assets. Traditional financial measures do not cover these intangible assets.

The balanced scorecard suggests that organizations should be viewed and measured from four different perspectives. These perspectives are as follows:

- **The business process perspective**—the internal business processes that are often classified as mission oriented and support oriented. Examples of this perspective include the length of time spent prospecting and the amount of rework required.
- **The customer perspective**—the level of customer focus and customer satisfaction. Examples of this perspective include the amount of time spent on customer calls and customer survey data.
- **The financial perspective**—the financial aspects of the organization. Examples of this perspective include financial ratios and various cash flow measures.
- **The learning and growth perspective**—includes employee training and organizational attitudes related to both employee and organizational improvement. Examples of this perspective include the amount of revenue that comes from new ideas and measures of the types and length of time spent training staff.

Using the balanced scorecard, companies create their own unique measures of these four aspects of the business. The specific measures that a company develops should reflect the specific drivers and strategy of the business.

Kaplan and Norton recommend a nine-step process for creating and implementing the balanced scorecard in an organization.

1. Perform an overall organizational assessment.
2. Identify strategic themes.
3. Define perspectives and strategic objectives.
4. Develop a strategy map.
5. Drive performance metrics.
6. Refine and prioritize strategic initiatives.
7. Automate and communicate.

8. Implement the balanced scorecard throughout the organization.

9. Collect data, evaluate, and revise.

There are many benefits and challenges to the balanced scorecard. The primary benefit is that it helps organizations translate strategy into action. By defining and communicating performance metrics related to the overall strategy of the company, the balanced scorecard makes the strategy come alive. It also enables employees at all levels of the organization to focus on important business drivers.

The main challenge of this system is that it can be difficult and time-consuming to implement. Kaplan and Norton originally estimated that it would take an organization a little more than 2 years to fully implement the system throughout the organization. Some organizations implement in less time and some require more time. The bottom line is that the balanced scorecard requires a sustained, long-term commitment at all levels in the organization for it to be effective.

—Joan P. Brannick

See also Measurement Scales; Performance Appraisal; Total Quality Management

FURTHER READING

Kaplan, R. S., & Norton, D. P. (1993, September). Putting the balanced scorecard to work. *Harvard Business Review, 71,* 134–147.

Kaplan, R. S., & Norton, D. P. (1996, January). Using the balanced scorecard as a strategic management system. *Harvard Business Review, 74,* 75–85.

Kaplan, R. S., & Norton, D. P. (2000, September). Having trouble with your strategy? Then map it. *Harvard Business Review, 78,* 167–176.

Kaplan, R. S., & Norton, D. P. (2004, February). Measuring the strategic readiness of intangible assets. *Harvard Business Review, 82,* 52–63.

Niven, P. R. (2002). *Balanced scorecard step-by-step: Maximizing performance and maintaining results.* New York: Wiley.

Ulrich, D., Zenger, J., & Smallwood, N. (1999). *Results-based leadership.* Boston: Harvard Business School Press.

BANDING

Banding refers to the procedure of grouping test scores into ranges and treating scores within a particular range as equivalent when making personnel decisions. After an organization collects test scores from candidates who applied for a job, a hiring decision must be made using these scores. There are a number of approaches for making these decisions. One common strategy is called *top-down selection:* Candidate scores are ranked from highest to lowest and organizations start at the top of the list by selecting the candidate with the highest score, then move to the person with the next highest score, and so on down the list. Another common strategy is the practice of setting cutoff scores. A cutoff score involves setting a passing score where candidates at or above this score are labeled as passing the test, whereas those below are labeled as failing. With a cutoff score those passing are treated as if they performed equally on the test. Banding is an alternative to top-down and cutoff score approaches.

Banding involves creating a defined range within which candidate scores are treated as being the same. This is similar to grouping scores into grades as done in many academic settings (e.g., a score between 90% and 100% is considered an A, a score between 80% and 89% is considered a B, etc.). The concept of banding is based on the idea that small differences between test scores may not translate into meaningful differences in expected job performance. For example, a candidate who scores 94% on a test may not perform noticeably better on the job than a candidate who scores 92%. This is because tests are not perfectly predictive of job performance and have varying degrees of measurement error. Banding is the idea of taking into account this imprecision by creating ranges within which test scores are treated as being the same. Thus for candidates who have scores that fall within the same band, the difference between their scores is viewed as meaningless in terms of predicting meaningful differences in job performance, and therefore the candidates are treated as if they scored equivalently on the test.

PURPOSE OF BANDING

One key question is, Why would an organization create bands within which candidate scores are considered equivalent? Critics have argued that banding results in a loss of information and has a negative impact on the utility or usefulness of a test. They state that a top-down approach has the highest utility. In response others have noted that although banding may

in some circumstances result in a loss of economic utility, this loss may be negligible and must be weighed against other compelling reasons for banding such as the need to increase workforce diversity.

Banding was first proposed as a method for reducing the adverse impact against protected groups (e.g., minorities, women) that is often associated with a top-down approach to selection test decision making. This is because Whites, on average, tend to outperform certain minorities on commonly used written multiple-choice selection tests measuring factors such as cognitive ability and job knowledge. Given this situation, Whites will be chosen at a substantially higher rate in comparison to members of these minority groups if a strict top-down rank order approach is used. Banding was suggested as a viable strategy for addressing this problem. Banding can reduce adverse impact because a band includes lower-scoring as well as higher-scoring individuals; thus when selection decisions are made regarding whom to choose from a band, other factors such as diversity can be taken into account. That is, if candidates that fall within a band are considered equal, an organization may consider the minority group membership of candidates when deciding whom to hire from a given band rather than just selecting the individual with the highest score. Banding allows an organization the flexibility to consider other factors such as diversity when making hiring decisions, whereas a top-down approach does not.

CREATING BANDS

Many different methods exist for developing bands. For example, expert or managerial judgments could be used to determine what range of scores on a test should be considered equivalent. Another viable approach is to use historical data on how candidates in the past performed on the test and subsequently on the job to determine what bands should be formed. An additional, yet controversial, method for creating bands is the concept of using indicators test reliability as a basis for creating bands. This approach uses statistical significance testing to determine what size the band should be so that test scores that fall within a band are not considered statistically different.

The most common version of this approach leverages a statistic known as the standard error of the difference (SED) to create bands. This SED procedure specifies a range of test scores that will be treated as statistically indistinguishable at some accepted level of confidence. That is, the bandwidth is a function of the standard error of measurement of the test and the desired level of confidence that scores within a band are not statistically different. This approach leverages the psychometric properties of the test in terms of its reliability to determine proper bandwidth. Critics of this approach state that the logic behind it is fatally flawed and that carrying it out to its conclusion would lead to random selection (i.e., selecting individuals completely at random rather than based on their test scores). However, proponents of this approach note that because selection tests are not perfectly reliable, the degree of unreliability should be taken into account when interpreting test scores. They further state that using indicators of unreliability is a more objective and appropriate way to create bands than doing so based on purely arbitrary decisions or solely relying on *expert* judgments.

TYPES OF BANDS

Bands can be either fixed or sliding. Fixed bands use the top score as the starting point, and the first band consists of all scores that fit within the range of the top score minus the bandwidth. For example, if the top score on a test was 96.0 and the bandwidth based on the SED approach was 5.2, the first band would range from 96.0 to 90.8. All scores that fell within this range would be considered part of band one and they would be treated as if they were equivalent. The second band would be the next highest score after band one minus the bandwidth. Therefore, for the example given earlier, the second band would range from 90.7 to 85.5. Additional bands would be created in a similar manner. With a fixed band approach, all individuals within a given band must be selected prior to moving to the next band. That is, band one needs to be completely exhausted before moving to band two.

Sliding bands also use the top score as an initial starting point, and the band is equal to this starting point minus the bandwidth. However, the difference with sliding bands is that when the top score is selected, a new band is formed using the next highest existing score in the band as the starting point. That is, when a top score in a band is chosen, the band slides down and is established using the next highest score as its anchor point. Using the previous example where the top score was 96.0 and the bandwidth was 5.2, individuals would be chosen from within this band until the top score is chosen, at which time the band

would slide down to be anchored on the next highest score. Thus if the individual with a score of 96.0 was chosen and the next highest score was 94.0, the new band would be set at 94.0 minus the bandwidth of 5.2 (i.e., a range of 94.0 to 88.8). Furthermore, when the current high score of 94.0 is chosen, the band would slide again and anchor on the next highest remaining score. The sliding band allows an organization to consider more people more quickly by moving down the rank order list more rapidly. Unlike with fixed bands, sliding bands do not require that a band be exhausted before moving down the list. Instead, when a top score is chosen, the band slides down and allows the organization to consider new individuals for selection.

EFFECTIVENESS AND LEGALITY OF BANDING

Research has shown that depending on varying circumstances, such as bandwidth size, banding can be used to reduce adverse impact. An organization can use banding procedures to group scores and then give preference to certain groups when selecting from a band as a means of increasing the diversity of its workforce. Opponents of banding note the loss in utility from not using a top-down approach, but proponents have responded by stating that the possible loss in economic utility is not substantial. One other key issue is whether banding is a legal practice. Most agree that although banding is legal, choosing individuals from a band based on protected group status (e.g., race, gender) could be problematic. The Civil Rights Act prohibits considering factors such as race and gender when making hiring decisions. Although this issue has not been fully resolved, recent court rulings have upheld the use of different types of banding. However, a review of these cases indicates that when protected group status is the only factor used to make choices from a band, it is less likely to be acceptable to the courts than when it is only one of many factors that are used.

—Harold W. Goldstein

See also Adverse Impact/Disparate Treatment/Discrimination at Work; Selection Strategies

FURTHER READING

Aguinis, H. (2004). *Test-score banding in human resource selection: Technical, legal, and societal issues.* Westport, CT: Praeger Publishers.

Campion, M. A., Outtz, J. L., Zedeck, S., Schmidt, F. L., Kehoe, J. F., Murphy, K. R., et al. (2001). The controversy over score banding in personnel selection: Answers to 10 key questions. *Personnel Psychology, 54,* 149–185.

Cascio, W. F., Outtz, J. L., Zedeck, S., & Goldstein I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4,* 233–264.

Henle, C. A. (2004). Case review of the legal status of banding. *Human Performance, 17,* 415–432.

BEHAVIORAL APPROACH TO LEADERSHIP

The behavioral approach to leadership involves attempts to measure the categories of behavior that are characteristic of effective leaders. Two research projects, one at Ohio State University and another at the University of Michigan, are most commonly associated with the behavioral approach to leadership. The results of both research programs suggested that the behavior of effective leaders could be classified into two general categories. The behavioral approach dominated leadership research throughout most of the 1950s and 1960s.

THE OHIO STATE STUDIES

Immediately following World War II, a group of scholars, including Carroll L. Shartle, John K. Hemphill, and Ralph M. Stogdill, conducted a series of investigations that became known as the Ohio State Leadership Studies. Rather than focusing on the traits or styles of effective leaders, as had been the focus of much early psychological research on leadership, these researchers studied the behaviors that leaders engaged in during the course of their interactions with followers. In a review of early leadership research, Stogdill (1963) declared that attempts to discover the traits shared by effective leaders had largely failed. This presumed failure, coupled with the rise of the behaviorist school of psychology, which emphasized behaviors rather than personality or mental processes, helped prompt the abandonment of trait-oriented leadership research and the rise of the behavioral approach.

Using detailed observations of leaders' behaviors, as well as reports from the leaders themselves and

from their subordinates, the Ohio State researchers accumulated a list of hundreds of leader behaviors. From these a list of 150 statements was derived that represented unique leader behaviors, such as "He assigns group members to particular tasks" and "He finds time to listen to group members." These 150 items composed the first form of the Leader Behavior Description Questionnaire (LBDQ). The LBDQ was administered to workers who rated how often their leaders engaged in each of the behaviors, using a five-point scale from *never* to *always.*

The responses to these items were subjected to factor analysis. The results suggested that the various leader behaviors clustered into one of two factors or categories: initiation of structure and consideration. Initiation of structure includes leader behaviors that define, organize, or structure the work situation. For example, clearly defining roles, assigning specific tasks, communicating work-related expectations, emphasis on meeting deadlines, maintaining standards of work performance, and making task-related decisions are all examples of initiation of structure behaviors. The orientation of these initiation of structure behaviors is focused primarily on the work task.

Consideration behaviors are those where leaders show concern for the feelings, attitudes, needs, and input of followers. They include the leader developing rapport with followers, treating them as equals, showing appreciation for their good work, demonstrating trust in followers, bolstering their self-esteem, and consulting with them about important decisions. The considerate leader is concerned with follower job satisfaction and with developing good interpersonal relationships with and among members of the work group.

The Ohio State researchers concluded that these two leadership behavior dimensions, initiation of structure and consideration, were not opposite ends of a continuum. They were independent of each other. In other words, both were independently related to effective leadership. They found that some effective leaders displayed high levels of initiation of structure behaviors, others engaged in high levels of consideration behaviors, and some displayed high levels of both. Only low incidences of both initiation of structure and consideration behaviors were associated with ineffective leadership.

The two dimensions of initiation of structure and consideration struck a responsive chord with leadership scholars, and a great deal of research followed. One line of research examined the robustness of the initiation of structure and consideration dimensions. Those results were generally supportive, suggesting that most leader behavior can indeed be grouped into one of the two general categories.

Research also refined the LBDQ. It was first reduced to 40 items, and a special version, the Supervisory Behavior Description Questionnaire, was constructed to measure the behavior of lower-level managers. A final revision yielded the LBDQ–Form XII, consisting of 10 items measuring initiation of structure and 10 items measuring consideration. The LBDQ-XII is the most widely used in research and is still readily available to scholars.

Additional research investigated the relationship between the two categories of leader behavior and work outcomes. For example, initiation of structure was found to correlate positively with effective group performance, but the relationship between initiation of structure and group member job satisfaction is less clear. There is some evidence for a positive relationship, but some conflicting evidence suggests a possible negative correlation between initiation of structure and job satisfaction, with a corresponding increase in employee turnover. Conversely, leader consideration was found to correlate positively with follower job satisfaction, but there have been inconsistent findings regarding work group performance. Correlations between leader consideration and performance have ranged from slightly positive to slightly negative.

These inconsistent results led researchers to conclude that the effectiveness of these broad categories of initiation of structure and consideration leader behaviors was likely dependent on contingencies in the leadership situation. Factors such as the type of work task, the structure of the work group and organization, the size of the group, and the level of the leader (e.g., executive versus middle manager versus frontline supervisor) can all influence how initiation of structure and consideration relate to key outcomes such as group performance and satisfaction.

THE UNIVERSITY OF MICHIGAN STUDIES

About the same time as the Ohio State studies, researchers at the University of Michigan, including Rensis Likert, Robert L. Kahn, Daniel Katz, Dorwin Cartwright, and others were also focusing on leader behaviors, studying leaders in several large, industrial organizations. They reached a conclusion similar to

the one reached by the Ohio State researchers. Leader behavior could indeed be clustered into two broad categories. The Michigan State researchers distinguished between task-oriented (also referred to as *production-oriented*) and relationship-oriented (also referred to as *employee-oriented*) leader behaviors.

Task-oriented leader behaviors tend to focus on performing the work group's job and are similar to initiation of structure behaviors. Task-oriented behaviors include setting clear work standards, directing followers' activities, instructing them on work procedures, and meeting production goals. Relationship-oriented behaviors focus more on employee well-being and allowing them to participate in decision-making processes, similar to consideration behaviors. The main difference between the Ohio State and the University of Michigan approaches was that the Michigan results suggested that relationship-oriented leader behaviors were more effective overall than task-oriented behaviors, but both types of leader behaviors were displayed by the most highly effective leaders. This makes intuitive sense considering research findings that suggest stronger connections between task-oriented leader behaviors and group performance and relationship-oriented behaviors and follower satisfaction, rather than vice versa. Therefore leaders who are both task and relationship oriented should turn out workers who are *both* productive and satisfied.

This notion influenced the development of the Leadership Grid, a leadership intervention program designed to foster both task- and relationship-focused leader behaviors. In the Leadership Grid, leaders are taught to be concerned with both production and people. Leaders who demonstrate both categories of leader behavior are seen as *team leaders,* whereas those who lack both are considered *impoverished.*

CONTRIBUTIONS AND LIMITATIONS OF THE BEHAVIORAL APPROACH

The main contribution of the behavioral approach to leadership is the explication of two very different forms of leader behavior: those that focus on the work task and those that focus on the follower. The fact that two independent lines of research arrived at the same two general categories suggests that these factors are clear and distinct.

The primary limitation of the behavioral approach was suggested by the research findings. How could such very different forms of leader behavior—focusing on the task, versus focusing on the people—both lead to effective leadership in some cases but not in others? The answer is that elements of the situation interact with styles of leader behavior to determine when the two categories of leader behavior might be effective and when they are not. This led to the development of contingency, or situational models, of leadership that examined the interaction between leader behavior and styles and variables in the situation that facilitate effective leadership. Although the situational theories of leadership go beyond the simple focus on leader behavior, most incorporate the results of the behavioral approach as an important element of their models.

—Ronald E. Riggio

See also Situational Approach to Leadership; Trait Approach to Leadership

FURTHER READING

Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications* (3rd ed.). New York: Free Press.

Blake, R. R., & McCanse, A. A. (1991). *Leadership dilemmas, grid solutions.* Houston, TX: Gulf.

Kahn, R., & Katz, D. (1960). Leadership practices in relation to productivity and morale. In D. Cartwright & A. Zander (Eds.), *Group dynamics: Research and theory* (2nd ed.). Elmsford, NY: Row, Peterson, & Co.

Kerr, S., & Schriesheim, C. A. (1974). Consideration, initiating structure, and organizational criteria: An update of Korman's 1966 review. *Personnel Psychology, 27,* 555–568.

Likert, R. (1961). *New patterns of management.* New York: McGraw-Hill.

Stogdill, R. M. (1963). *Manual for the Leader Behavior Description Questionnaire–Form XII.* Columbus: Ohio State University, Bureau of Business Research.

Stogdill, R. M., & Coons, A. E. (Eds.). (1957). *Leader behavior: Its description and measurement.* Columbus: Ohio State University, Bureau of Business Research.

BENCHMARKING

Organizations use a variety of measurements to evaluate business performance, such as revenue, stock price, voluntary attrition, or employee attitude survey results. Comparing these measures to relevant

benchmarks provides decision makers with a standard that can be used to interpret the organization's standing and draw meaningful conclusions. The standard, target, or benchmark can be derived from internal organizational data or from data external to the organization. Benchmarking databases are similar to normative data used in clinical psychological testing to establish parameters for normal and abnormal results. Although benchmarking commonly uses numeric data for comparisons, nonnumeric benchmarking is also used to aid decision making, in areas such as strategic organizational direction, or in processes such as supply chain or marketing. The benefits, caveats, and sources of benchmarking are addressed in the following text.

BENEFITS

The benefits of benchmarks are to provide an empirically substantiated target figure that is more realistic and has more credibility and weight than one determined subjectively, such as *gut feeling.* Targets created in an internal vacuum may result in setting goals that are neither challenging enough nor attainable. Research has shown that these types of goals are de-motivating. Although benchmarks based on internal organizational data can be constructed, external benchmarks, especially when competitors are involved, have gravitas that usually gets the attention of executives who make strategic decisions regarding an organization's future direction and can serve to inspire positive changes. In addition, benchmarking allows *best practices* to be identified: approaches yielding results surpassing all others.

CAVEATS

There are a number of caveats regarding external benchmarking. Some benchmark databases are composed of samples of convenience that may contain comparison groups that are neither relevant nor equivalent, thus making differences between an organization's scores and the benchmark of little or no value. Other benchmark databases may be of questionable quality. These types of poorly constructed benchmarks again can result in setting de-motivating goals that are either unrealistic or not challenging enough. Similar to the need for norm groups that are representative of the population of interest in interpreting scores on clinical psychology tests (e.g., matched on salient demographics; excluding those with impairments), comparable

organizations, such as same industry and similar size, are best for benchmarking purposes in a business setting. However, it is important to keep in mind that even organizations in quite different industries may be similar on other dimensions such as competition to recruit the same top employee talent. In this case obtaining external benchmarks on such things as workplace climate provides a context for evaluating an organization's position as an employer of choice.

Economic and cultural differences are also important to consider and control for to develop appropriate business benchmarks. For example, comparing business results in countries in emerging economies to results in countries with more established economies is not a useful comparison. An additional example comes from employee opinion research where it is widely known that employees in some countries typically have more positive responses compared with employees in other countries. If these responses are pooled across all countries, an organization with representation in countries with typically less favorable responses will be compared with a database skewed in a more positive direction.

In addition, comparability of question translations across companies that contribute data to a benchmark database needs to be considered when evaluating benchmarks for global employee opinion surveys. For example, two different question translations may result in different interpretations of the question, thus producing poor-quality benchmarks. Some consortia attempt to solve this problem by establishing common translations that organizations must use to submit data to the benchmarking database.

INTERNAL BENCHMARKING

Internal comparisons can avoid some criticisms applied to external benchmarking. Types of internal benchmarking include tracking trends over time, polling executives to set goals, or identifying perceived gaps between executives' expectations and the current state in an organization. However, in the absence of external values, it can be difficult to determine reasonable targets for an outcome or strategic direction. That is, internal improvements over time may not be enough if results remain below par compared with competitors. Further, internal benchmarks across different units within a single organization can promote unhealthy internal competition versus all internal efforts being directed at external competitors.

Internal improvements also may reach a ceiling, a numeric level that is typically not exceeded. For example, in workplace climate research the best achievable score for employee satisfaction with compensation is routinely much lower than the best achievable score for satisfaction with organizational teamwork. Without external benchmarks, these *ceiling* differences would be unknown, and unrealistic targets for improvement could be set.

BENCHMARKING SOURCES

Sources of external benchmarks include nonprofit consortia. Consortia typically collect a fee, have rigorous standards for company membership and data contributions, and a hire a third party vendor to collect data and provide reports to members. Consortia have additional benefits in terms of cross-company information sharing, networking, and standardization of instruments for collecting the data submitted to the benchmarking database. Well-known consortia exist in the life insurance and banking or financial industries. Others focus on specific metrics such as employee opinion survey research.

Consultants may also provide benchmarking data to clients using their client base as the source. Benchmarking data may also be available via public information sources, such as financials for publicly traded companies.

The list of organizations included in the database—their size, the number of data points from each organization, the countries from which the data originate, and the time frame in which the data were collected—are all important questions to ask the provider when evaluating benchmark data.

SUMMARY

External and internal benchmarking are extremely valuable organizational tools that potentially provide appropriate targets and direction for actions that can contribute to greater success in an organization. Recognizing the limitations and applications of benchmarking and benchmarking sources will avoid obtaining inaccurate data that may lead to misinformed decision making and ill-directed corporate strategizing.

—*Sara P. Weiner*

See also Measurement Scales; Performance Appraisal

FURTHER READING

Campbell, A. (1999, March–April). Tailored, not benchmarked: A fresh look at corporate planning. *Harvard Business Review,* 41–50.

Johnson, R. H. (1996). Life in the consortium: The Mayflower Group. In A. I. Kraut (Ed.), *Organizational surveys* (pp. 285–309). San Francisco: Jossey-Bass.

Morris, G. W., & LoVerde, M. A. (1993). Consortium surveys. In P. Rosenfeld, J. Edwards, & M. D. Thomas (Eds.), *Improving organizational surveys: New directions, methods, and applications* (pp. 122–142). Newbury Park, CA: Sage.

Rogelberg, S. G., Church, A. H., Waclawski, J., & Stanton, J. M. (2002). Organizational survey research. In S. G. Rogelberg (Ed.), *Industrial and organizational psychology* (pp. 141–160). Malden, MA: Blackwell.

BIG FIVE TAXONOMY OF PERSONALITY

Personality traits are characteristic behaviors, thoughts, and feelings of an individual that tend to occur across diverse situations and are relatively stable over time. Given this broad definition, literally thousands of personality traits can be identified. For the better part of 100 years, personality researchers have attempted to create a standard taxonomy, or organizing structure, of personality traits. Although some disagreement remains, the Big Five taxonomy is currently the dominant perspective on the organization of personality traits. The Big Five traits are identified in the following text, and trait descriptive terms are provided for each:

1. **Neuroticism:** Anxious, temperamental, nervous, moody versus confident, relaxed, unexcitable

2. **Extraversion:** Sociable, energetic, active, assertive versus shy, reserved, withdrawn, unadventurous

3. **Openness:** Intellectual, innovative, artistic, complex versus unimaginative, simple, unsophisticated

4. **Agreeableness:** Trusting, trustful, helpful, generous versus cold, harsh, rude, unsympathetic

5. **Conscientiousness:** Organized, neat, thorough, systematic, efficient versus careless, undependable, haphazard, sloppy

The term *Big Five* was coined by Lewis R. Goldberg in 1981 and was meant to signify that these traits are

broad in nature. Generally, the Big Five trait taxonomy is conceptualized as hierarchical, such that the Big Five traits are the broadest level. Within each of the Big Five traits, narrower trait dimensions can be defined, representing the second level of the hierarchy. As one progresses to lower points in the hierarchy, increasingly narrow trait dimensions can be identified. The lowest level of the taxonomy consists of specific behaviors.

ORIGINS OF THE BIG FIVE

Although human curiosity and examination of personality traits dates back to the ancient Greeks, the history of the Big Five begins with the work of Gordon W. Allport and Henry S. Odbert based on the lexical hypothesis. The lexical hypothesis suggests that important aspects of human behavior will be encoded into language; and the more important an aspect is, the more likely it will be encoded as a single word. Based on this hypothesis, Allport and Odbert turned to the dictionary to identify the basic elements of personality. They identified almost 18,000 personality related words, and organized these terms into four categories:

1. personal traits,
2. temporary moods,
3. evaluative terms (such as excellent or irritating), and
4. miscellaneous.

Although Allport and Odbert stopped with the identification of these personality descriptive terms, Raymond B. Cattell sought to bring order to them. Cattell began his work with the set of nearly 4,500 words Allport and Odbert placed in the personal trait category. As this set of terms was far too large to investigate empirically, Cattell conceptually combined the terms into 171 clusters. Still too numerous to work with given that his computations needed to be done by hand, he eventually worked his way down to a set of 35 clusters. He was then able to collect data on these clusters and conduct a factor analysis. Finally, he arrived at a set of 12 factors, but many believe that he overfactored the data.

Based on Cattell's work, two factor-analytic studies provided a foundation for what would eventually become the Big Five. Using 22 of Cattell's 35 clusters, Donald W. Fiske in 1949 and Ernest C. Tupes and Raymond E. Christal (1961/1992) found five

similar factors when scores from the 22 clusters were factor analyzed. The Tupes and Christal findings were particularly interesting in that they found the five factors within each of eight samples that differed in many ways: education (high school graduates, college students, graduate students), type of rating (self-ratings, peer ratings), and among the peer ratings, length of acquaintanceship (from 3 days to 1 year or more). The five factors identified by Fiske and by Tupes and Christal were defined in a manner that is similar to the way in which the Big Five are defined today.

Recognizing some of the limitations in Cattell's conceptual sorting of the trait terms, Warren T. Norman went back to the beginning and developed a new list of trait descriptive terms from the dictionary. Norman, like Allport and Odbert before him, sorted his set of terms into broad categories and focused his work on those terms that fell into the category he labeled biophysical traits. After doing considerable work to reduce the set of terms in this category to roughly 1,550 terms, he set out to organize them. First, the terms were sorted into the endpoints of the five factors identified by Tupes and Christal, giving him 10 groups of words. He then sorted each of the 10 groups of words, which resulted in 75 groups of words. A factor analysis of scores on these groups produced the expected five factors.

To this point, much of the research on the five factors had been directly related to the initial work of Cattell. Recognizing this fact, Goldberg (1990) conducted studies on trait terms that were common in the English language, finding the same five-factor structure. Given that these words were selected on the basis of common usage and not on the variables identified by Cattell, these studies demonstrated that the five factors were general and not specific to Cattell's variables. The Big Five was born.

To date, considerable research has been conducted to establish the Big Five. Numerous questionnaire measures of the Big Five traits have been developed, suggesting that the factors are not unique to the study of trait descriptive terms. The five factors have been found in a wide variety of cultures from across the globe in both adjective and questionnaire measures; and evidence suggests that they are, at least in part, heritable.

LINGERING TAXONOMIC ISSUES

Although the Big Five is the dominant perspective on the organization of personality traits, there remain

differences of opinion regarding some aspects of the taxonomy. Hans Eysenck (1992) and Auke Tellegen, for example, have argued that the highest level of the taxonomy should be represented by three rather than five traits. Eysenck has vigorously defended his position that the highest level of the taxonomy should be represented by the traits of extraversion, neuroticism, and psychoticism, a perspective that some have referred to as the *Even Bigger Three.* Although extraversion and neuroticism are defined by Eysenck in a manner that is consistent with the Big Five, he argues that psychoticism is made up of lower levels of conscientiousness and agreeableness. Tellegen has taken a position similar to Eysenck's, arguing that the three traits of positive emotionality (extraversion and part of conscientiousness), negative emotionality (neuroticism and low agreeableness), and constraint (part of conscientiousness and low openness) should dominate the highest levels of the trait taxonomy.

Some debate also remains about the names and definitions of some of the Big Five traits themselves. For example, the agreeableness dimension has also been referred to as love, likability, and nurturance, each of which conveys a somewhat different interpretation. Oliver John has argued, in fact, that the term *agreeableness* is somewhat misleading, suggesting a submissive nature that would actually be located at the lower end of the extraversion trait. Although the term *conscientiousness* seems to be well accepted at this point in time, this trait has also been referred to by various authors as dependability, work, will to achieve, responsibility, and constraint. Perhaps the most controversy, however, has surrounded the nature of the openness dimension. In addition to openness, this dimension has been referred to as culture, intellect, and intellectance. The controversy stems from the apparent incorporation of aspects of intelligence into the factor. For example, in Goldberg's work, the term *intelligent* was consistently an indicator of this dimension. Some researchers have been highly critical of the association of this dimension with intelligence, fearing that the dimension will be considered synonymous with intelligence as measured by IQ tests when, in fact, the dimension is much broader, encompassing artistic and creative aspects, a willingness to try new things, and a sense of open-mindedness.

It seems that much of the controversy surrounding the naming of the five dimensions is a result of their broad nature. Some clarity might be brought to the

issue if there were to be consensus regarding the next lowest level of the trait hierarchy. Scant work, however, has been done to identify and define the traits at the level below the five dimensions. There is some consensus among industrial/organizational (I/O) researchers interested in personality that the trait of conscientiousness can be broken down into two dimensions of achievement striving and dependability. Also, Robert and Joyce Hogan have argued that extraversion can be split into sociability and ambition. It seems clear that research focusing explicitly on this level of the hierarchy is warranted.

One problem with establishing the lower levels of the trait hierarchy is that the hierarchy is likely to be reticulated. That is, many lower-level traits are liable to relate to more than one trait at the higher levels. Using studies of adjectives as a source of examples, some researchers have associated warmth with extraversion whereas others have associated it with agreeableness. Likewise, the characteristic of impulsiveness has been associated with neuroticism, extraversion, and conscientiousness by various researchers. These cross-associations of traits at one level with traits at higher levels will make the process of achieving consensus at levels of traits below the Big Five difficult, but it would seem to be a worthwhile endeavor.

It is important to recognize that the Big Five taxonomy is simply descriptive and is not a theory. As such, it does not explain why people behave in the ways they do; it is only a system for classifying behavioral tendencies. Although many have criticized the Big Five because it is not theoretical, others have argued that the taxonomy is necessary before theory can be developed. To this end, Paul Costa and Robert McCrea have proposed a five-factor theory of personality. Although the theory is broad in scope, at its core it suggests that the Big Five are a result of biological processes and influence people's characteristic adaptations—the ways they think, feel, and behave in their unique environments.

THE BIG FIVE AND INDUSTRIAL/ ORGANIZATIONAL PSYCHOLOGY

The emergence of the Big Five through the 1980s was a tremendous benefit to both I/O-related research and the application of personality testing in organizational contexts. Although multitrait personality inventories began to emerge in the 1930s, the use of personality testing in applied settings was largely haphazard and

not theoretically grounded before the emergence of the Big Five. Reviews of the criterion-related validities of personality scales conducted in the 1950s suggested little support for using personality tests for predicting job performance. As noted by Robert Guion and Richard Gottier in 1965, the field seemed to be dominated by a *broadside* approach where every available personality test score was correlated with all available performance measures. Although many of the observed correlations were small, Guion and Gottier asserted that many of these would be expected, based on theory, to be small.

The emergence of the Big Five allowed researchers and practitioners to select traits (and scales representing those traits) based on a conceptual mapping of the traits to the performance dimension. As a result, numerous meta-analyses on the relationships between personality test scores and measures of work performance have resulted in positive findings regarding the criterion-related validities of personality tests. These meta-analyses have generally shown that conscientiousness is related to almost all job-related criteria (i.e., performance, training, attendance, etc.) across almost all jobs. Other Big Five dimensions have also proven important predictors but not as universally as conscientiousness. For example, extraversion has been shown to be related to performance in managerial and sales jobs, and openness has been related to training performance.

Having established the usefulness of personality testing, many researchers are exploring factors that may strengthen or weaken the personality–performance relationship. In 1993, for example, Murray Barrick and Michael Mount examined the extent to which the degree of autonomy given to employees would moderate the relationship between conscientiousness and job performance. They found that in autonomous situations (i.e., where workers had more control over their activities), the relationship between conscientiousness and job performance was stronger than in situations where workers were given less autonomy.

One contentious issue has been whether practitioners should use broad or narrow traits to predict performance; that is, whether to focus on the Big Five or on more narrow traits at some point lower in the hierarchy. Although authors on both sides of this issue have made potent arguments for their perspectives, it would appear that the solution is to attempt to match the breadth of the predictor with that of the criterion.

When predicting broad criteria, it appears optimal to use broad traits such as the Big Five. In contrast, when more narrow criteria are of interest, narrower trait constructs are preferred.

CONCLUSIONS

The Big Five trait taxonomy is the dominant organizing structure for personality traits. Although the traits emerged from the lexical approach to personality, the structure is found with questionnaire measures and identified in cultures around the world. The impact of the Big Five on the role of personality in I/O research and application has been immense, allowing for theoretically guided predictor-criterion mapping. At present, there is a great deal of interest in personality within the field of I/O psychology, an interest in no small part a result of the Big Five taxonomy of personality traits.

—Eric D. Heggestad

See also Factor Analysis; Individual Differences; Personality; Personality Assessment

FURTHER READING

Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationship between the Big Five personality dimensions and job performance. *Journal of Applied Psychology, 78,* 111–118.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117,* 187–215.

Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13,* 653–665.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences, 13,* 667–673.

Goldberg, L. R. (1990). An alternative description of personality: The Big-Five factor structure. *Journal of Personality and Social Psychology, 59,* 1216–1229.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135–164.

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.

Tupes, E. C., & Christal, R. E. (1961/1992). Recurrent personality factors based on trait ratings (ASD-TR-61-97).

Lackland Air Force Base, TX: Aeronautical Systems Division, Personnel Laboratory. (Reprinted in 1992 in the *Journal of Personality, 60,* 225–251)

BIOGRAPHICAL DATA

Biographical data, or biodata, are measures of key aspects of individuals' life experiences intended to predict job applicants' future performance in organizations, whether that performance is task-specific job performance, teamwork, or shoplifting. Although biodata can be developed to measure a wide array of experiences and psychological constructs, the fundamental and general premises underlying the predictive power of biodata measures are that

- individuals in free societies shape their life experiences, and they also are shaped by them;
- this process of reciprocal influence between personality and situations occurs over a large time span; and therefore,
- measures of past experience should predict future work behavior, especially given a relatively unconstrained environment where employees' typical performance can be wide-ranging.

In light of these premises, items on a biodata measure can be relatively personality oriented or covert in nature (e.g., "To what extent does your happiness depend on how things are going at work?"), or they can be relatively situation oriented and overt in nature (e.g., "Approximately how many books have you read in the past three months?"). In either case responding involves some cognitive processing where test takers are required to recall and summarize information, the accuracy of which depends on the accuracy of prior perception and storage, and in many cases the saliency or recency of the event.

Although biodata can vary widely in their content and constructs measured and can be scored in different ways, they have consistently demonstrated moderate to high levels of validity across job types (approximately .30); they also demonstrate incremental validity beyond ability and personality measures in predicting performance. Constituent biodata items either explicitly or implicitly reflect constructs such as ability, personality, motivation, interpersonal skills, and interests. They can be relatively pure measures of these constructs; however, biodata items that ask test takers about their experiences may be related to a combination of constructs, not just one. Analyses of the latter type of items may result in a weak general factor in a factor analysis or a low alpha reliability coefficient. Both test–retest reliability and alpha reliability should be considered when attempting to measure the stability of scores on biodata measures.

ITEM ATTRIBUTES

An outline of 10 major attributes of biodata items was proposed by F. A. Mael and is as follows:

1. Historical versus hypothetical (past behaviors versus predicted behaviors in the future, or behaviors in *what-if* scenarios)

2. External versus internal (behaviors versus attitudes)

3. Objective versus subjective (observable or countable events versus self-perceptions)

4. Firsthand versus secondhand (self-descriptions versus how people would say others describe them)

5. Discrete versus summative (single events versus averaging over a period of time)

6. Verifiable versus nonverifiable

7. Controllable versus noncontrollable (circumstances that could or could not be influenced by a decision)

8. Equal access versus unequal access (access to opportunities with respect to the group being tested)

9. Job relevant versus nonjob relevant

10. Noninvasive versus invasive

SCORING METHODS

Historically, biodata measures have developed out of a tradition of strong empiricism, and therefore a wide variety of scoring methods have been proposed. The criterion-keying approach involves taking individuals' responses to a given biodata item and calculating the mean criterion score or the criterion-related validity for each response option. This is done for each item, and these values are used as item response weights for scoring purposes. Weights may be rationally adjusted when nonlinear patterns in relatively continuous response options are found or when some weights are based on small sample sizes. A similar approach to criterion keying can be taken when keying biodata items not to criteria but rather to personality or

temperament measures. This is a particularly interesting approach in keying a set of objective or verifiable biodata items, which tend to be less susceptible to faking but often are harder to assign to single psychological constructs. (Even if such keying is not done, it remains helpful to place the biodata measure within a nomological net of cognitive and noncognitive constructs.) When biodata items can be assigned to constructs in a relatively straightforward manner, such as by developing item content around constructs or through an a priori or post hoc subject matter expert (SME) item-sorting procedure, a straightforward scoring of each item along a single underlying continuum may be possible as is done with traditional Likert-scale self-report measures of personality.

Configural scoring is an entirely different approach to scoring biodata items, because it involves grouping individuals into representative profiles of biodata scores. Subgroups are defined, both conceptually and methodologically, as internally consistent yet externally distinct, similar to the interpretation of statistically significant group differences in the analysis of variance. Individuals are often assigned to subgroups based on their similarity to a subgroup mean, such as in *k*-means analysis; or sometimes a set of data is aggregated until the appropriate balance between parsimony and descriptiveness is reached, such as in Ward's method. Subgroup profiles may then be labeled (e.g., *goal-oriented social leaders* or *emotional underachievers*) and then related to relevant external criteria, or profiles of criteria, for purposes such as personnel selection and placement; or subgroup profiles can be used in their own right for training and development.

Two general points regarding the scoring of biodata items are worth noting. First, any appropriate scoring method should be informed by both rational and empirical approaches. Being purely rational or theory based ignores important empirical data that could serve to revise the theoretical underpinnings that generated the biodata items in the first place—or at least it could revise subsequent item-development rules. Conversely, being purely empirical in the absence of a theoretical or conceptual rationale would impede, if not preclude, appropriate item development, item revision, and score use and interpretation. Second, item-scoring methods that are developed on one sample should be cross-validated on an independent sample, such as a holdout sample from the original data set or an entirely different sample. Doing so helps ensure that the

features of the model are generalizable and not sample specific; for example, cross-validation can ensure that increased validity, reduction of group mean differences, or a *cleaned up* exploratory factor analysis result achieved in one sample by selectively reweighting or removing biodata items can then be achieved in an independent sample using the same subset of items, so that the original results (in large part, at least) cannot be attributed to capitalization on chance. The same concern applies to regression models, where least-squares regression weights may capitalize on chance and thus artificially inflate validity. In this case, cross-validation formulas can be applied to the whole sample, to estimate what the shrinkage in validity would be should those weights be applied to an independent sample of the same size.

RACE DIFFERENCES

Because biodata items vary widely in content, no general statement about race differences can be made that is of any use. At a more specific level, however, biodata containing culturally relevant content have demonstrated Black–White subgroup differences in terms of differential item functioning (DIF). Black–White differences in biodata have also been found in the domain of swimming proficiency. Other race differences are likely when the biodata measures are aligned with constructs where it is known that race differences exist, such as general cognitive ability or certain personality traits.

APPLICANT REACTIONS

Meta-analysis indicates that studies using biodata measures generally show a favorability (i.e., job relevance and fairness) rating at about the midpoint of the scale, with measures such as interviews, résumés, and cognitive ability tests showing greater favorability and personal contacts and integrity tests showing less favorability. Although the meta-analytic mean across studies is stable, nontrivial variability in favorability ratings across studies exists; this is likely because of the variety of biodata measures that can be developed. This highlights a consistent theme in the research literature: Biodata measures tend to be viewed more favorably when they are perceived as relevant to the job at hand and part of a fair personnel selection system.

—Frederick L. Oswald

See also Individual Assessment; Individual Differences; Person–Job Fit; Prescreening Assessment Methods for Personnel Selection

FURTHER READING

Dean, M. A., & Russell, C. J. (2005). An examination of biodata theory-based constructs in a field context. *International Journal of Selection and Assessment, 2,* 139–149.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology, 44,* 763–927.

Mount, M. K., Witt, L. A., & Barrick, M. R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology, 53,* 299–323.

Oswald, F. L., Schmitt, N., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college performance. *Journal of Applied Psychology, 89,* 187–207.

Oullette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin, 124,* 54–74.

Reiter-Palmon, R., & Connelly, M. S. (2000). Item selection counts: A comparison of empirical key and rational scale validities in theory-based and non-theory-based item pools. *Journal of Applied Psychology, 85,* 143–151.

BONA FIDE OCCUPATIONAL QUALIFICATIONS

United States federal fair employment laws generally prohibit discrimination in employment on the basis of certain *protected characteristics,* including race, color, religion, sex, national origin, age, and disability. However, the fair employment laws permit employers to discriminate based on a protected characteristic in rare situations where the characteristic is considered a bona fide occupational qualification (BFOQ) for the job in question.

The BFOQ defense is potentially available in those Title VII cases where it has been established, and not merely alleged, that an employer's employment policy intentionally discriminated on the basis of religion, sex, or national origin. The BFOQ defense does not apply to discrimination based on race or color. It is also potentially available in cases involving employer policies that have been shown to intentionally discriminate the basis on age (Age Discrimination in Employment Act) or disability (Americans With Disabilities Act). Where successfully asserted, the BFOQ defense allows employers to treat job applicants or employees differently depending on their protected class status (religion, sex, national origin, age, disability), making permissible conduct that would otherwise be considered illegal discrimination. For example, although Title VII generally prohibits discrimination against job applicants based on their sex, if it is established that being male is a BFOQ for the job in question, the employer may lawfully refuse to consider women for the job. However, it is important to understand that the BFOQ defense is narrowly written and extremely narrowly construed by the courts and that employers asserting the defense have the burden of proving that its stringent requirements (discussed in the following text) are met. As a result, the BFOQ defense is available in relatively few situations.

REQUIRED ELEMENTS OF THE BFOQ DEFENSE

To establish a BFOQ, an employer must meet two requirements. First, the employer must prove that a strong, direct relationship exists between the protected characteristic in question (e.g., sex) and an employee's ability to perform one or more functions of the job in question. Second, the employer must prove that the functions of the job to which the protected characteristic is directly related are important functions that go to the *essence* or *central mission* of the employer's business operation.

The Direct Relationship Requirement

The direct relationship requirement must be met by showing either that *all or substantially all* members of the group that is being excluded based on a protected characteristic cannot perform the functions of the job, or that it is *impossible or highly impractical* to determine on an individual basis whether members of the excluded group can perform the functions of the job. For example, an employer seeking to justify a sex-based BFOQ that would allow it to hire only men must show either that all or substantially all females are unable to perform the functions of the job, or that it would be impossible or highly impractical to assess female applicants' qualifications to perform the job

functions on an individual basis, for example, through the use selection tests.

It is clear that the *all or substantially all* standard can be met without proof that 100% of the excluded class cannot perform the functions of the job in question. However, it is also clear that the employer must produce credible evidence of a strong relationship between the protected characteristic and the ability to perform the job. Relying on stereotypes about the abilities or disabilities of women, older workers, and so on is insufficient. It is also not enough to merely show that members of the excluded group, on average, tend not to perform the job as well. Further, given the vast and growing array of selection tools that are available to assess job applicant qualifications on an individual basis (e.g., assessments of physical strength, motor skills, cognitive ability), it is extremely difficult for employers to successfully argue that they should be able to use a protected characteristic as a general hiring or promotion criterion because it is *impossible* or *highly impractical* to assess applicants' qualifications on a more individualized basis.

Essence of the Business Requirement

It is not enough to show a direct relationship between the protected characteristic in question and a job function that is only incidentally or marginally related to the employer's business operations. The protected characteristic must be directly related to the ability to perform one or more important job functions that are closely associated with the fundamental purpose(s) of the employer's business. This means that to determine whether an asserted BFOQ is justified, the court must determine the primary purpose(s) or *essence* of the business operation in which the job is embedded.

Cases considering whether safety concerns support the BFOQ defense illustrate how the *essence of the business* requirement affects whether the BFOQ defense is available to employers. Safety concerns may be the basis for a BFOQ but only if the safety concern is indispensable to the particular business at issue. For example, the safety of inmates was found to be a legitimate basis for a sex-based BFOQ applied to the prison guard position, because the safety of inmates goes to the core of a prisons guard's job performance and the essence of the business in which prisons are engaged. In contrast, when considering whether to exclude female employees of childbearing age from jobs involving exposure to toxic material, the Supreme Court held that concerns about the safety of female employees' unborn children may not be the basis for a BFOQ because the essence of the employer's business was manufacturing batteries, and the fetuses of female employees were neither customers nor third parties for whom safety is essential to the business of manufacturing batteries.

General Guidance

Although the availability of the BFOQ defense is determined on a case-by-case basis, and there is some variation in how lower courts interpret and apply the Supreme Court's rulings in this area, useful guidance for assessing the availability of the BFOQ defense can be provided based on court cases, the legislative history, and EEOC (Equal Employment Opportunity Commission) guidelines. In addition to safety concerns, BFOQs have been recognized based on privacy concerns where, again, those concerns relate to the essence of the employer's business, such as sex-based BFOQs for bathroom attendant and masseur positions. The BFOQ defense has also been recognized when viewed as necessary to ensure the genuineness or authenticity of an employer's business operations. Examples of *authenticity* BFOQs include the use of male and female actors to play male and female roles in theater productions and a restaurant hiring only ethnic chefs where a primary goal of the employer is to maintain an authentic ethnic atmosphere.

The courts have uniformly refused to accept discriminatory customer preferences or biases as a basis for a BFOQ, usually noting that these biases are the type of discrimination that fair employment laws such as Title VII were intended to eliminate. For example, courts have refused to accept the preferences of male customers as a legitimate basis for a BFOQ allowing the hiring of only female flight attendants and have rejected the argument that being male was a BFOQ for an overseas assignment because customers and associates in other countries preferred to do business with men.

Finally, it is well settled that the BFOQ defense cannot be based merely on the incremental or extra cost associated with hiring one protected group versus another. Thus, for example, an employer cannot exclude women from a certain position merely because of concerns that allowing women to occupy the position

(i.e., not restricting the position to men) may result in greater health- or liability-related costs for the employer.

—*Mark V. Roehling*

See also Age Discrimination in Employment Act; Americans With Disabilities Act; Civil Rights Act of 1964, Civil Rights Act of 1991

FURTHER READING

Berman, J. B. (2000). Defining the "essence of the business": An analysis of Title VII's privacy BFOQ after Johnson Controls. *University of Chicago Law Review, 67*, 749–775.

Kapczynski, A. (2003). Same-sex privacy and the limits of antidiscrimination law. *The Yale Law Journal, 112,* 1257–1294.

Lindeman, B., & Grossman, P. (1997). *Employment discrimination law* (3rd ed.). Washington, DC: The Bureau of National Affairs.

McGowan, S. M. (2003). The bona fide body: Title VII's last bastion of intentional sex discrimination. *Columbia Journal of Gender and Law, 12,* 77–127.

BOREDOM AT WORK

Feeling bored at work is a common complaint; a large percentage of employees feel bored at least occasionally and some feel bored much of the time. Boredom has not been studied extensively, but it has attracted some attention from scholars in diverse disciplines including human factors engineering, psychiatry, sociology, education, criminology, and industrial psychology.

DEFINITIONS OF BOREDOM

Most scholars would agree that boredom is an emotion. It is an unpleasant transient state in which individuals feel an extreme lack of interest in their current activity. Bored individuals find it difficult to keep their attention focused on work and may feel that time is passing very slowly. Boredom is usually accompanied by feelings of restlessness, irritability, and desire to escape or change the situation to a more interesting activity. Boredom has been described as the opposite of enthusiasm or *flow.*

Boredom is also sometimes conceptualized as a personality trait, and some individuals are more likely to experience boredom than others. Scores on the boredom proneness scale are related to measures of state boredom, impulsiveness, sensation seeking, depression, negative affect, aggression, hostility, self-reported physical and psychological symptoms, and job dissatisfaction. The remainder of this entry will focus on boredom as a transient state experienced while working.

CONSEQUENCES AND CAUSES OF BOREDOM AT WORK

The consequences of boredom are thought to be largely negative. Boredom at work has been associated with absence, dissatisfaction, accidents, reduced performance on vigilance tasks, performance variability, horseplay, and sabotage. However, it has been suggested that boredom has the potential to stimulate creativity and organizational citizenship behaviors in some cases.

There are many likely causes of boredom at work. These include aspects of work tasks, aspects of the surrounding work environment, and interactions of the task and performer.

Work Tasks

As an emotion, boredom depends on an appraisal of a situation by the performer. Thus boredom does not automatically reside in characteristics of work tasks but in how these tasks are appraised by the individual performing them. Nevertheless, there are types of tasks that are likely experienced as boring by most people. What makes a task seem boring is at least partly the opposite of what makes it interesting or intrinsically motivating. Simple, repetitive tasks that require little thought or judgment, such as some assembly line tasks, are likely to be experienced as boring. Once learned, these tasks require little conscious attention, provide little mental stimulation, and may prohibit incumbents from engaging in other forms of self-entertainment while working.

Another type of work that is often experienced as boring includes vigilance, inspection, checking, and driving tasks. These tasks require sustained and careful attention. However, they provide little variety or stimulation in return. This makes it difficult to sustain

attention and perform with high reliability over long periods of time.

A final category of work situation that is described as boring is having nothing to do. Some jobs do not contain enough tasks to keep incumbents occupied for the time they are required to remain at work. Other jobs are dependent on intermittent or less than completely predictable demand for services, such as checkout or help desk staff. When demand is low, there may be little to do but wait around in readiness to provide a service.

Work Environment

Compulsion and interruptions can also contribute to feelings of boredom while working, regardless of characteristics of the main work task. Individuals report feeling bored when they are compelled to perform tasks in set ways, in set places, and at set times. Lack of self-direction, autonomy, and personal causality are known to undermine intrinsic interest in work tasks.

Individuals may infer that they are bored when they experience problems holding their attention on a work task. Some research has suggested that low-level distractions and interruptions in the workplace can make maintaining attentional focus difficult, thus contributing to the experience of boredom. Interruptions can also stem from internal sources. Personal concerns may produce intrusive thoughts that distract an incumbent from a work task so it appears uninteresting.

Interactions of Task and Performer

Some authors attribute boredom largely to lack of personal meaning in an activity. Individuals are bored when they perform a task that lacks relevance for them. Simple repetitive tasks often fit this description, as might any required task when something else is more important or has greater meaning to the performer at that moment. Individuals are also bored when tasks are too difficult for their skills. Tasks may be varied and complex, but the performer lacks the expertise to extract meaning from the complexity. An example is listening to a lecture that is too advanced for a person's level of understanding.

REDUCING BOREDOM

Both individuals and organizations may act to reduce boredom. Bored employees adopt a number of strategies to alleviate their unpleasant feelings. Sometimes it is possible to force attention on to the task and eventually become absorbed in it. Another option is to engage in *subsidiary behaviors* to provide additional stimulation while performing the boring task. For example, a worker may fidget, talk to others, daydream, listen to music, or invent varied ways to execute the task. If the task does not require full attention, these strategies may reduce boredom without compromising performance. Performance on vigilance tasks, however, will often suffer when subsidiary behaviors are performed. Alternatively, individuals may escape or avoid boring situations altogether by finding different work or nonwork tasks to do: engaging in counterproductive work behaviors such as horseplay or sabotage, taking breaks, being absent, or quitting the job.

Organizations may adopt job rotation or job enrichment and redesign to increase the variety and challenge in employees' tasks and thus reduce boredom. Frequent feedback, goal setting, and performance-contingent pay can make simple tasks more meaningful and therefore less boring. Although there is no research evidence yet, team-based work systems also might reduce boredom. Allowing social contact between workers and permitting other forms of concurrent self-entertainment can help reduce boredom on simple repetitive tasks. Because boredom occurs when skills are either too high or too low for task demands, creating an appropriate match between demands and skills through selection, training, and job design should minimize boredom.

—Cynthia D. Fisher

See also Intrinsic and Extrinsic Work Motivation; Job Characteristics Theory; Job Satisfaction; Role Overload and Underload

FURTHER READING

Barbalet, J. M. (1999). Boredom and social meaning. *British Journal of Sociology, 50,* 631–646.

Conrad, P. (1997). It's boring: Notes on the meanings of boredom in everyday life. *Qualitative Sociology, 20,* 465–475.

Damrad-Frye, R., & Laird, J. D. (1989). The experience of boredom: The role of self-perception of attention. *Journal of Personality and Social Psychology, 57,* 315–320.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment, 50,* 4–17.

Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations, 46,* 395–417.

Fisher, C. D. (1998). Effects of external and internal interruptions on boredom at work: Two studies. *Journal of Organizational Behavior, 19,* 503–522.

Smith, R. P. (1981). Boredom: A review. *Human Factors, 23,* 329–340.

Vodanovich, S. J. (2003). Psychometric measures of boredom: A review of the literature. *The Journal of Psychology, 137,* 569–595.