

Validating the Selection Process



Jack Kelly Clark

“A couple of years ago we started experimenting with a new hiring procedure for our pruning crews. I feel the only fair way to hire pruners is through a practical test. We don’t have the problem any more of hiring people who claim to know how to prune only to find after they are on the job that they don’t know. I think 10 to 15 years from now a pruning test will be the standard for the industry.”¹

**Vineyard Manager
San Joaquín Valley, California**

After recently sharing my enthusiasm for testing in New Zealand, I was asked: *Who tests the test?* The answer: the process of test validation. *Validity* is a measure of the effectiveness of a given approach. A selection process is only valid if it helps a farm employer increase the chances of hiring the right person for the job. It is possible to evaluate hiring decisions in terms of such valued outcomes as high picking speed, low absenteeism, or a good safety record. A selection process is not valid on its own, but rather, relative to a

specific purpose. For example, a test that effectively predicts the work of strawberry pickers may be useless in the selection of a capable crew foreman.

A critical component of validity is *reliability*. Validity embodies not only what positive outcomes a selection approach may predict, but also how *consistently* (i.e., reliably) it does so. In this chapter we will (1) review ways of improving the consistency or reliability of the selection process; (2) discuss two methods for measuring validity; and (3) present cases that illustrate these

methods. First, however, let's consider a legal issue that is closely connected to validity: employment discrimination.

AVOIDING DISCRIMINATION CHARGES

It is illegal—and a poor business practice—to discriminate on the basis of such *protected characteristics* as sex, race, color, age (40 or older), national origin, and a host of others. In terms of discrimination one can distinguish—to use the language of the courts—between (1) disparate *treatment* and (2) adverse *impact*. Outright discrimination, or disparate treatment, involves treating people differently on the basis of a protected classification.

Examples of such illegal personnel decisions are disqualifying all women from arc-welding jobs on the assumption that they cannot operate the equipment, or hiring field workers only if they were born in Mexico.

Practices that appear unbiased on the surface may also be illegal if they yield discriminatory results—that is, if they have adverse impact. For instance, requiring a high school diploma for tractor drivers might eliminate more minority applicants from job

consideration. If not shown to be related to job performance, this requirement is illegal even though there appears to be nothing discriminatory about the practice—or perhaps even about the intent. In another example, a policy that requires all applicants to lift 125-pound sacks—regardless of whether they will be hired as calf feeders, pruners, office clerks, or strawberry pickers—might have an adverse impact on women.

Clearly, it is legal to refuse employment to unqualified—or less qualified—applicants regardless of their age, sex, national origin, disability or the like. You are not required to hire unqualified workers. Employers, however, may be expected to show that the selection process is job related and useful.²

An employer can give applicants a milking dexterity test and hire only those who do well. If a greater proportion of women passed the test, more women would be hired—on the basis of their test performance, not of their gender.

If women consistently did better than men, however, the farmer could not summarily reject future male applicants without testing them. Such a practice would constitute disparate treatment. In general, the greater the adverse impact, the greater the burden of proof on employers to defend the validity of their selection process if it is challenged.

The Americans with Disabilities Act is likely to cause an increase in the number of job opportunities for disabled individuals. A systematic selection approach, one where applicants have the chance to demonstrate their skills, is more likely to help you meet the requirements of this law. Instead of treating people with disabilities differently, where one might make assumptions about who can or cannot do a job, all applicants have the same opportunity to demonstrate their abilities. In some instances, applicants with disabilities may ask for specific accommodations.

Research has shown that people tend to make unfounded assumptions about others based on such factors as height and attractiveness. Obtaining more

An employer can give applicants a milking dexterity test and hire only those who do well. If a greater proportion of women passed the test, more women would be hired—on the basis of their test performance, not of their gender.



SIDEBAR 3-1**Suggestions for Interaction with the Disabled³**

(1) Speak directly to the person rather than to a companion of the disabled.

(2) Focus on the person's eyes, not the disability. (This is especially so when speaking to someone who is severely disfigured.)

(3) Be patient. (If a person has a speaking disability, formulated thoughts may not be expressed easily. Also, be patient with the mentally retarded and those whose disabilities may reduce activity or speed of communication.)

(4) Remember, a disabled person has feelings and aspirations like everyone else (even though muscles, hearing, or eyes may not work as well).

(5) Refrain from hasty assumptions that uncoordinated movement or slurred speech are the result of intoxication.

(6) Use slower speed but a normal tone of voice to speak with someone with a hearing impairment (no need to shout).

(7) Do not cover your mouth when talking to someone with a hearing impairment (they may read lips).

(8) Write down the message if needed, when communicating with the hearing impaired.

(9) Announce your general intentions with the visually impaired (introduce yourself, announce your departure).

(10) Avoid gestures when giving instructions to the visually impaired.

(11) Offer to cut food when meals are involved; for those with muscular disabilities, have food pre-cut in the kitchen; tell those with visual disabilities where their food, utensils, and so on are placed, in terms of a clock (e.g., your milk is at 12 o'clock, knife at three o'clock).

(12) Avoid panicking if an individual has a seizure (you cannot prevent or shorten it). Instead, (a) protect the victim from dangerous objects she may come in contact with; (b) avoid putting anything between the victim's teeth; (c) turn the victim's head to the side when he relaxes; and (d) allow the victim to stay where she is until consciousness is regained.

(13) If you do offer help, make sure it is completed (e.g., don't abandon a blind person before he knows his exact location).

(14) Remember, the person with the impairment is the expert on how he can be helped.

detailed information about an applicant's merits can often help employers overcome stereotypes and avoid discriminatory decisions. For instance, I know of a dedicated journeyman welder who can out-weld just about anyone, despite his missing the better part of an arm. Suggestions for interaction with the disabled are offered in Sidebar 3-1. A well-designed selection approach can help farmers make both legal and effective hiring decisions.

IMPROVING SELECTION RELIABILITY

For a selection process to be valid, it must also be reliable. That means the process must measure what it is

designed to measure, and do so consistently over time. For instance, how consistently can a Brix refractometer gauge sugar content in table grapes? How reliable is a scale when measuring the weight of a calf? And how often does an employee

Obtaining more detailed information about an applicant's merits can often help employers overcome stereotypes and avoid discriminatory decisions.





For a selection process to be valid, it must also be reliable. That means the process must measure what it is designed to measure and do so consistently over time.

selection process result in hiring effective workers?

Reliability is measured in terms of both (1) selection scores and (2) on-the-job performance ratings. If either measure is unreliable (i.e., selection testing or on-the-job evaluation), the process will not appear to be valid. No matter how consistently workers pick apples, for instance, if an apple-picking test yields different results every time it is given to the same person, the lack of test consistency will result in low validity for the overall procedure. More often, however, it is the on-the-job performance measures that lack consistency. Performance appraisals are often heavily influenced by the subjective evaluation of a supervisor (Chapter 6).

Reliability may be improved by ensuring that (1) the questions and activities associated with the selection process reflect the job accurately; and

(2) raters reduce biases and inconsistencies in evaluating workers' performance.⁴

Avoiding content errors

Content errors occur when different applicants face unequal appraisal situations, such as different sets of questions requiring dissimilar skills, knowledge, or abilities. One applicant for the job of vineyard manager, for example, might be asked about eutypa and mildew and another questioned on phylloxera and grapeleaf skeletonizer.

As applicants may do better with one set of questions than the other, all should be presented with approximately the same items. Content errors may be reduced by carefully identifying the most important skill requirements for that job. Some flexibility is needed to explore specific areas of different applicants' qualifications, but the greater the variance in the questions presented, the greater the potential for error.

Hiring decisions should not be based on partial results. It can be a mistake to get overly enthusiastic about one candidate before all the results are in, just as it is a mistake to eliminate candidates too freely. It is not unusual, for instance, for a candidate to shine during the interview process but do poorly in the practical test—or vice versa.

Reducing rater inconsistency

Rater inconsistency accounts for a large share of the total unreliability of a measure. Objective indicators are more likely to be reliable than subjective ones, but even they are not totally free from scorer reliability errors (e.g., recording inaccuracies).

One manager felt his seven supervisors knew exactly what to look for in pruning a young orchard. After a little prodding, the manager agreed to a trial. The seven supervisors and a couple of managers discussed—and later set forth to judge—pruning quality. Four trees were designated for evaluation. Supervisors who thought the tree in the first row was the best pruned were asked to raise their hands. Two went up.

Others thought it was the worst. The same procedure was followed with subsequent trees, with similar results.

In another situation, four well-established grape growers and two viticulture farm advisors participated in a pruning quality study. As in the preceding situation, quality factors were first discussed. Raters then went out and scored ten marked vines, each pruned by a different worker. As soon as a rater finished and turned in his results, to his surprise he was quietly asked to go right back and rate the identical vines again. The raters' ability to evaluate the vines consistently varied considerably. It is clearly difficult for each rater to be consistent in his own ratings, and it is even more difficult to achieve consistency or high reliability among different raters.

Having applicants rate everything twice is very telling, especially in those instances where individuals are required to constantly make judgments about which animals need treatment, picking and packing decisions, or evaluating the quality of employees' work. Applicants may be asked to rate 100 or more items before repeating the exercise with the same items (but with a fresh evaluation sheet). An excellent tool to gauge applicant skill is the GageRR statistical tool (Chapter 11).

Here are eight areas where you can reduce rating errors:

1. *Present consistent challenges to applicants.* You can draw up a list of job-related questions and situations for interviews, practical tests, and reference checks (see Chapter 2).

Rules and time limits should be applied in a like manner for all candidates. If one foreman allows more time or gives different instructions to applicants taking a test, resulting scores may differ between equally qualified persons.

2. *Use simple rating scales.* The broader the rating scale, the finer the distinctions among performance levels. A scale of 0 to 3 is probably easier to work with consistently than a scale of 1 to 10 (see Figure 3-1). I find the following way to think about these numbers helpful: a 0 means the

Vineyard Pruning-Quality Scorecard			
Quality factor	Rating	Weight	Score
Fruiting wood selection		x4	
Spur placement		x3	
Spur number		x2	
Spur length		x2	
Closeness of cut		x2	
Angle of cut on spur		x1	
Distance of cut from bud		x1	
Removal of suckers		x1	
Total:			
<small>Rate each category from a <i>three</i> (superior) to a <i>zero</i> (intolerable). Then multiply <i>rating</i> by the <i>weight</i> to obtain the <i>score</i>. Determine what the mistake tolerance for each quality factor will be, ahead of time, for a given sample of vines evaluated.</small>			

FIGURE 3-1
Pruning Score Card.

individual did not perform this task; a 1 means that the task was performed poorly; a 2 means the individual was able perform the task; and finally, a 3 means the person excelled in performing the task. With these basic anchors, it is then possible to add a plus or a minus to distinguish performance ability even further. For instance, a task may be rated as a 2+ or a 3-. Such a scale then provides 9 different points (0, -1, 1, 1+, -2, 2, 2+, -3, 3).

3. *Know the purpose of each challenge.* If it is difficult to articulate either the reason for including a question or what a good response to it would be, perhaps the item should be rephrased or eliminated.

4. *Reduce rater bias.* Raters need training, practice opportunities, and performance feedback. Utilize only tested and proven judges who are accurate and consistent. Provide clear scoring guidelines. When possible, it helps to break down potentially subjective ratings into objective components. (Chapters 6 and 11 speak extensively on some of these rating skills.)



It is possible for an instrument to measure consistently yet still be useless for predicting success on the job.

Consider the farmer who hires cherry-pickers on the basis of their understanding of picking quality. Once on the job, these workers may be paid solely on the basis of speed.

5. Decide between multiple and single raters. In Chapter 11, we look at the advantages of utilizing few but very reliable and accurate raters for quality control work.

When multiple raters are used, these may function in either a single or a sequential approach; that is, applicants may face one or several raters at a time (e.g., in the interview process). One advantage of having multiple raters function at one time is the common ground on which to discuss applicant performance. Raters may need to defend the logic of their conclusions. Improper questioning and abuse of power may be discouraged.

In order to preserve independent perceptions, multiple raters should not share their evaluations until all candidates have performed. Most raters are influenced, if not swayed, by the opinions of others—especially when there are different hierarchical levels. Avoiding premature discussion takes self-discipline.

One disadvantage of reviewing candidates at the end is that perceptions are no longer fresh in each rater's mind. Time for raters to take adequate notes between candidates is therefore crucial.

Sometimes raters seem more concerned with justifying their stand than with hiring the best person for the job. This may become apparent when a

rater finds only good things to say about one candidate and bad things about the rest. A skillful moderator, who is less invested in the position being filled, may help. This facilitator can help draw out shy raters and help manage disagreement among more aggressive ones. Positive and negative qualities about each candidate can be jotted down or displayed where all can see. Finally, participants can disclose their rankings for further discussion.

6. Pretest each step of the selection process for time requirements and clarity. Trying out interviews and tests in advance helps fine-tune contents and determine time limits. A trusted employee or neighbor who goes through the selection steps can advise you on modifications that improve clarity or reasonableness. Moreover, the results from a pretest can be used to help train raters to evaluate applicant performance.

Not infrequently, a query “matures” during successive interviews. As they repeatedly ask a question, interviewers sometimes realize that another question was really intended. The selection process is fairer to all if the correction is made before the actual applicants are involved.

7. Pay close attention to the applicant. Carefully evaluating candidate performance takes concentration and good listening skills, so as to help raters avoid premature judgments. If as an interviewer you find yourself speaking more than listening, something is amiss. Effective interviewing requires (1) encouraging the applicant to speak by being attentive; and (2) maintaining concentration on the here-and-now. Because interviews can be such a mental drain, it is useful to build a break between each of them.

8. Avoid math and recording errors. Checking rating computations twice helps avoid errors. On one farm, foremen are asked to conduct and rate portions of a practical test. To simplify their task, however, the adding of scores—and factoring of weights—takes place back in the office.

We have said that it is possible for an instrument to measure consistently

yet still be useless for predicting success on the job. Consider the farmer who hires cherry-pickers on the basis of their understanding of picking quality. Once on the job, these workers may be paid solely on the basis of speed. The motivation for people to perform during the application process and in the course of the job might be quite different. There can still be a benefit to a selection approach that measures performance in a different job environment. Even when hiring for an hourly wage crew, for instance, a pruning test under piece rate conditions may be used to eliminate workers whose speed or quality are below a cutoff standard.

MEETING VALIDITY REQUIREMENTS

Two important means of establishing the validity of a selection instrument are the *statistical* and the *content* methods. A related consideration is “*face validity*”—though not really a validation strategy, it reflects how effective a test *appears* to applicants and judges (if it is ever contested in court). Ideally, a selection process is validated through multiple strategies. Regardless of which approach a farm employer uses, a rigorous analysis of the job to be filled is a prerequisite (Chapter 2).

The statistical strategy

A statistical strategy (the technical term is *criterion-oriented* validity) shows the relationship between the test and job performance. An inference is made through statistics, usually a *correlation coefficient* (a statistic that can be used to show how closely related two sets of data are, see Sidebar 3–2).

For example, a fruit grower might want to determine how valid—as a predictor of grafting ability—is a manual dexterity simulation in which farm workers have to quickly arrange wooden pegs in a box. If a substantial statistical relationship exists between performance on the test and in the field, the grower might want to use the simulation to hire grafters—who will never deal with wooden pegs in the real

job. Such an approach would be particularly useful when selecting foreign employees in their nations of residence.

The content-oriented strategy

In a content-oriented strategy, the content of the job is clearly mirrored in the selection process. This approach is useful to the degree that the selection process and the job are related. Thus, it makes sense for a herd manager who performs artificial insemination (AI) to be checked for AI skills, for a farm supervisor to role play a disciplinary process, and so on. The pitfall of this method is that people tend to be examined only in those areas that are easiest to measure. If important skills for the job are not tested, the approach is likely to be ineffective.

“Face validity” describes what a selection process appears to measure on the surface. For instance, candidates for a foreman position will readily see the connection between questions based on labor laws and the job.



SIDEBAR 3–2

Correlation Coefficients can be used to Gauge Reliability or Validity

The statistic essentially measures the extent to which two variables are linearly related. You cannot assume a cause-and-effect relationship just because of a high correlation. Factors may be related without one causing the other. Many computer programs can be used to quickly compute the *correlation coefficient* used in the statistical approach.

Correlations may range from -1 through 0 to a +1. A positive correlation indicates that applicants who do well on a test would do well on the job; those who do poorly on the test would do poorly on the job. A negative correlation indicates that applicants who do well on a test would do poorly on the job; those who do poorly on the test would do well on the job. A correlation coefficient score close to “0” would indicate the test and performance are not related. Expect correlation coefficients that measure reliability to be higher than those that convey validity (see table below, with subjective meanings for reliability and validity coefficients for a test consisting of about 20 to 40 applicants).

A related factor is that of *statistical significance*. Statistical significance answers the question, “Are these two factors related by chance?” The fewer

the number of pairs compared, the higher the correlation coefficient required to show significance. Software packages make it clear as to whether results yield statistical significance. Below I indicate what the correlation coefficients can mean (in evaluating the *strength* of a negative correlation coefficient using the table below, ignore the negative sign. For instance, instead of a -0.56, just read 0.56):

Meaning of Reliability Scores

<i>Correlation coefficient</i>	<i>Subjective Meaning</i>
r = .70 or greater	Somewhat acceptable
r = .80 or greater	Good
r = .90 or greater	Excellent

Meaning of Validity Scores

<i>Correlation coefficient</i>	<i>Subjective Meaning</i>
r = .40 or greater	Somewhat acceptable
r = .50 or greater	Good
r = .60 or greater	Excellent

When conducting tests it is informative to do two shorter tests so they can be checked for test-retest reliability. Sometimes applicants or other test takers worry that the process may be used to lower piece-rate wages (Chapter 10). When looking at the actual data, and fitting a regression line, it is possible to see when a few or many applicants perform unreliably.

Face validity

“Face validity” refers to what a selection process (or individual instrument) appears to measure on the surface. For instance, candidates for a farm human resource manager position will readily see the connection between questions based on agricultural labor law and the job. Although face validity is not a type of validation strategy, it is usually vital that a selection approach appear to be valid, especially to the applicant. A farmer wanting to test for a herdsman’s knowledge of math should

use test problems involving dairy matters, rather than questions using apples and oranges. The skills could be determined by either approach, but applicants often resent being asked questions that they feel are not related to the prospective job.

Face validity is a desirable attribute of a selection process. Not only does it contribute toward a *realistic job preview*, it also helps eliminate negative feelings about the process. Furthermore, anyone conducting a legal review is more likely to rule in favor of selection procedures appearing relevant.



SELECTION CASE STUDIES: PERFORMANCE DIFFERENCES

The following case studies should illustrate the practical application of statistical and content-oriented validation strategies.

Statistical strategy: testing of vineyard pruners⁵

Can a test—when workers know they are being tested—reliably predict on-the-job performance of vineyard pruners paid on a piece rate? Three hundred pruners—four groups on three farms—participated in a statistical-type study to help answer this question.

(Even though the emphasis of this test was on statistical evaluation, it clearly would also qualify as a content-oriented test: workers had to perform the same tasks during the test as they would on the real job.)

Selection test data. Workers were tested twice, each pruning period lasting 46 minutes. Pruners were told to work as fast as they could yet still maintain quality. A comparison of the results between the first and second test periods showed high worker consistency. There was a broad range of scores among workers: in one group, for instance, the slowest worker pruned just 3 vines in the time it took the fastest to prune 24. No relationship was found between

A statistical validation strategy shows the relationship between the test and job performance. For example, a 46 minute vineyard pruning test was shown to be a good predictor of worker performance on the job.

speed and quality, however. Some fast and some slow pruners did better-quality work than others.

Job performance data. On-the-job performance data was obtained from each farm's payroll records for two randomly selected days and two randomly selected grape varieties. To avoid influencing supervisors or crews in any way, on-the-job data was examined *after* the pruning season was over. Workers who had pruned quickly on one day tended to have pruned quickly on the other. Likewise, slow workers were consistently slow.

Validity. Significant valid relationships were found between the test and on-the-job performance measures. That is, workers who did well on the test tended to be the ones who did well on the job. The test was a good predictor of worker performance on the job. Similar results were obtained with hand-harvested tomato picking.⁶

Some may argue that it matters little if one hires effective workers as all are paid on a piece rate basis anyway. Some of the money farmers save as a result of hiring fewer, more competent employees

includes: (1) reducing the number of supervisors needed, (2) reducing fixed costs spent per worker regardless of how effective the worker is (e.g., vacation, training, insurance) and (3) establishing a reasonable piece rate. If some workers are very slow, the piece rate will need to be raised for all workers for these to be able to make a reasonable (or even a minimum) wage.

When there are few good applicants, employers may be tempted to lower the selection standards, but this only decreases the validity of the test.

Content strategy: secretarial selection

Our second case study illustrates a content-oriented validation strategy—used to hire a secretary to assist in my work for the University of California. Specific job requirements were identified.⁷ In developing a testing strategy, particular attention was paid to artistic layout and secretarial skills that would be needed on a day-to-day basis.

An advertisement specifying qualifications—including a minimum typing speed of 60 words per minute (WPM) and artistic ability—ran twice in

In a content-oriented strategy, the content of the job is clearly mirrored in the selection process. Thus, it makes sense for a herdsman who performs artificial insemination (AI) to be checked for AI skills, for a farm clerk-typist to be given a typing test, and so on.



Jack Kelly Clark

the local paper. Other recruitment efforts were made at a nearby college.

Of the 108 complete applications received, only a few reported typing speeds below 60 WPM. These were eliminated from consideration. All other applicants were invited to demonstrate their artistic layout ability. The quality of the artwork varied considerably among applicants, and was evaluated by three raters. The 25 applicants who performed at a satisfactory or better level were scheduled to move on to the next hurdle.

What applicants claimed they could type was at variance with their test scores (Figure 3–2). The average claimed typing speed was 65 WPM; the average tested speed about 44 WPM. The discrepancy between claimed and actual typing speeds was large (perhaps our test was more difficult than standard typing tests). More importantly, the test showed that some typists claiming higher ability than others, ended up typing slower. While there was an applicant claiming very fast speeds, and indeed she was swift, one could place little confidence on what applicants said they could type.

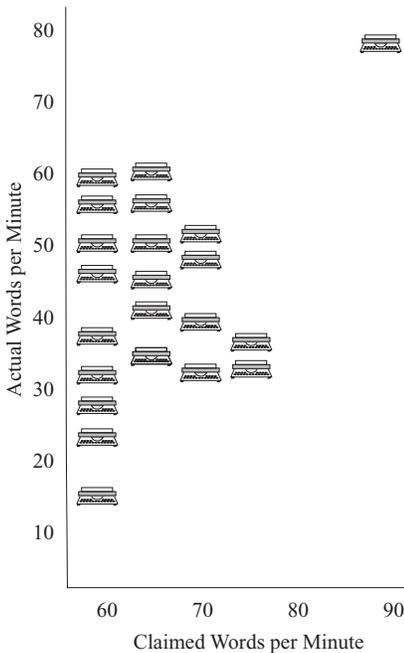


FIGURE 3–2
Secretarial Typing Speeds.

As a non-native English speaker, I still have some difficulties with sentence construction. For instance, I need to be reminded that I do not “get on my car” as I “get on my horse” (there is no such distinction in Spanish). We designed an appropriate spelling, grammar, and punctuation test. Applicants were provided a dictionary and asked to retype a letter and make necessary corrections. There was plenty of time allowed to complete the exercise.

Applicants ranged from those who found and corrected every mistake in the original letter (even some we did not know were there), to those who took correctly spelled words and misspelled them. Eight persons qualified for a final interview; three of these showed the most potential; one was selected unanimously by a five-person panel.

This content-oriented study also had “face validity” because the test was directly related to the performance required on the job. The selection process revealed the differences among more than 100 applicants. Had applications been taken at face value and the apparent top candidates interviewed, it is likely that a much less qualified candidate would have emerged. Moreover, the excellent applicant who was hired would normally not even have been interviewed: she had less secretarial experience than many others.

Content strategy: mechanic selection

Another content-oriented validation strategy was the selection of a farm mechanic at a large dairy operation. The farm owner set up about half a dozen stations where applicants were judged on their ability to troubleshoot farm equipment that did not start. Individuals were judged not only by their success, but by the systematic troubleshooting approach they took, as well as by the tools they selected to work on each project. Applicants had twenty minutes to work on each station. Additional testing involved looking up parts in a catalogue and ordering these on the phone, and demonstrating proper lifting techniques in order to avoid back injury.



This content-oriented study had “face validity” because the test was directly related to the performance required on the job. Had a secretarial applicant been selected without the benefit of a practical test, it is likely that a much less qualified candidate would have emerged.

This content-oriented study included a sampling of the duties of a mechanic and also had “face validity.”

SUMMARY

Agricultural managers interested in cultivating worker productivity can begin with the selection process. Any tool that attempts to assess an applicant’s knowledge, skill, ability, education, or even personality can itself be evaluated by how consistent (i.e., how reliable) it is and by how well it predicts the results it is intended to measure (i.e., how valid).

Improving the validity of a selection approach entails designing job-related questions or tests, applying them consistently to all applicants, and eliminating rater bias and error.

A content-oriented selection strategy is one in which the content of the job is clearly reproduced in the selection process. For example, applicants for an equipment operator position should be asked to demonstrate their tractor-driving skills, ability to set up a planter or cultivator, and other related tasks. A statistical strategy, on the other hand, studies the relationship between a test and actual job performance. A test may be useful even if it does not seem relevant at first glance. For instance, high performance on a dexterity test using tweezers may turn out to be a good indicator of grafting skill.

The validity of a specific selection instrument can be established by statistical or content-oriented strategies. Ensuring face validity will enhance applicants’ acceptance of the process. The more valid the selection instrument, the better chances a farmer has of hiring the right person for the job—and of successfully defending that choice if legally challenged.

A thorough employee selection approach brings out the differences among applicants’ abilities for specific jobs. Farmers should not depend too heavily on applicant self-appraisal to make their staffing choices. In the long run, a better selection process can help farmers hire workers who will be more productive, have fewer absences and

accidents, and stay longer with the organization.

CHAPTER 3 REFERENCES

1. Billikopf, G. E., and Sandoval, L. (1991). *A Systematic Approach to Employee Selection*. Video. This comment was made in the 1980s by a grapegrower who adopted testing for his vineyard pruners. Sadly his prediction, that job sample testing would become the industry norm, has not been fulfilled. Those employers who do use testing, however, have a great competitive advantage over those who do not.
2. Uniform Guidelines on Employee Selection Procedures. (1978). *Federal Register* Vol.43-166. Aug. 25. See also Vol. 44-43 (1979) and Vol.45-87 (1980). <http://www.uniformguidelines.com/uniformguidelines.html>.
3. For Those Who Serve the Public Face to Face. Glendale Partnership Committee for the International Year of Disabled Persons (1981). Reprinted by the Employment Development Department, State of California, Oct. 1990, along with comments from Charles Wall, Americans with Disabilities Act, Agricultural Personnel Management Association’s 11th Annual Forum, Modesto, California, March 7, 1991.
4. Anastasi, A. (1982). *Psychological Testing* (5th ed.) (p. 120). New York: Macmillan.
5. Billikopf, G. E. (1988). Predicting Vineyard Pruner Performance, *California Agriculture* (Vol. 42, No. 2) (pp. 13-14).
6. Billikopf, G. E. (1987). Testing to Predict Tomato Harvest Worker Performance, *California Agriculture* (Vol. 41, Nos. 5 and 6) (pp. 16-17).
7. Billikopf, G. E. (1988). *Agricultural Employment Testing: Opportunities for Increased Worker Performance*, Giannini Foundation Special Report No. 88-1. (pp. 17-18).

CHAPTER 3: ADDITIONAL RESOURCES

- (1) *Testing and Assessment: An Employer’s Guide to Good Practices*, http://www.cnr.berkeley.edu/ucce50/ag-labor/7labor/test_validity.pdf, U.S. Department of Labor Employment and Training Administration (1999) (80 pages).