Determining Optimum Acceptance Sample Size—A Second Look

Prepared By
Peter A. Kopac[1], Dooyong Cho[2], and Fazil T. Najafi[3]

(1) Corresponding Author, 9027 Gavelwood Court, Springfield, VA 22153
Phone: 703-451-1619
e-mail: pkopac@hotmail.com

(2) Department of Civil and Coastal Engineering,
University of Florida, 345 Weil Hall, Gainesville, FL 32611-6580
Phone: 352-215-5213
e-mail: joecdy@ufl.edu

(3) Department of Civil and Coastal Engineering, University of Florida,
345 Weil Hall, Gainesville, FL 32611-6580
Phone: 352-392-9537 ext. 1493
e-mail: fnaja@ce.ufl.edu

Number of Words:
5950 + 6 figures = 7450 words

Prepared for presentation and publication at the
90th Annual Meeting of the Transportation Research Board, January 2011

May 2010

Determining Optimum Acceptance Sample Size—A Second Look

ABSTRACT

What is the optimum acceptance sample size (n)? Recent research by Gharaibeh et al. found that, when incoming lots are anticipated to be high quality, the optimum n is relatively small ($n = 3$ for many acceptance quality characteristics). The finding was received with some skepticism, and many questions were raised regarding the optimization model and underlying assumptions. This paper summarizes follow-up research that more closely examined the optimization model elements and assumptions in an attempt to confirm or refute Gharaibeh's finding. With a somewhat modified model and assumptions based where possible on actual data, the conclusions reached by Gharaibeh stay basically the same and several new conclusions are drawn. Some cautions are presented regarding the state highway agency use of small sample sizes. Agencies are encouraged to perform their own optimization calculations. By doing so, they will gain a better understanding of their acceptance plan systems and associated costs, and have greater confidence in applying economic decision analysis principles to minimize expected costs and optimize statistical acceptance risks.

Key words: Sample Size, Acceptance Plans, Quality Assurance, Specifications.

Determining Optimum Acceptance Sample Size—A Second Look

The term "sample size" refers to the number of test results, *n*, obtained from a unit (i.e., a lot) of materials or construction. In practice, many state highway agencies divide their lots into five sublots and obtain one test result from each sublot; thus *n* is often 5. The objective of this paper is to report on an investigation of the optimum acceptance sample size *n*.

Up until recently, there had been limited research to determine the optimum *n* for *highway construction acceptance plans*. More research had been done for *quality control plans*. One such effort developed a cost optimization computer program (COSTOP1) to determine optimum quality control sampling frequencies; it concluded that higher frequencies of quality control tests than were commonly used (in 1985) would be cost-effective (*1*). With respect to quality acceptance, however, the recent research by Gharaibeh et al. (*2*) concluded that acceptance sample size *n* need not be large when incoming quality levels are satisfactory. For percent within limits (PWL) acceptance systems, Gharaibeh found *n* = 3 to be the optimum sample size for many acceptance quality characteristics (AQCs).

Gharaibeh's finding by no means contradicts the COSTOP1 research. Rather, it adds to our understanding of the relation between quality control testing and acceptance testing—the more testing done for the former, the less testing needed for the latter. Nonetheless, Gharaibeh's finding was quite unexpected as U.S. highway construction quality assurance (QA) practitioners have for years assumed that more acceptance testing is needed and would be cost-effective.

BACKGROUND

When highway construction QA specifications were first being developed in the 60's and 70's, the issue of acceptance sample size was approached several ways.

1. One way was to use the standard error of the mean to approximate the "required" sample size. As the curve in Figure 1 shows, there is a point where the normalized standard error of the mean, $\sigma_M = \sigma / \sqrt{n}$, has stabilized sufficiently such that any further increase in *n* has a negligible effect on σ_M. The exact point where this occurs is subject to interpretation of "sufficiently" and "negligible," but most agree it occurs at a relatively large sample size, i.e., *n* = 20 or more.

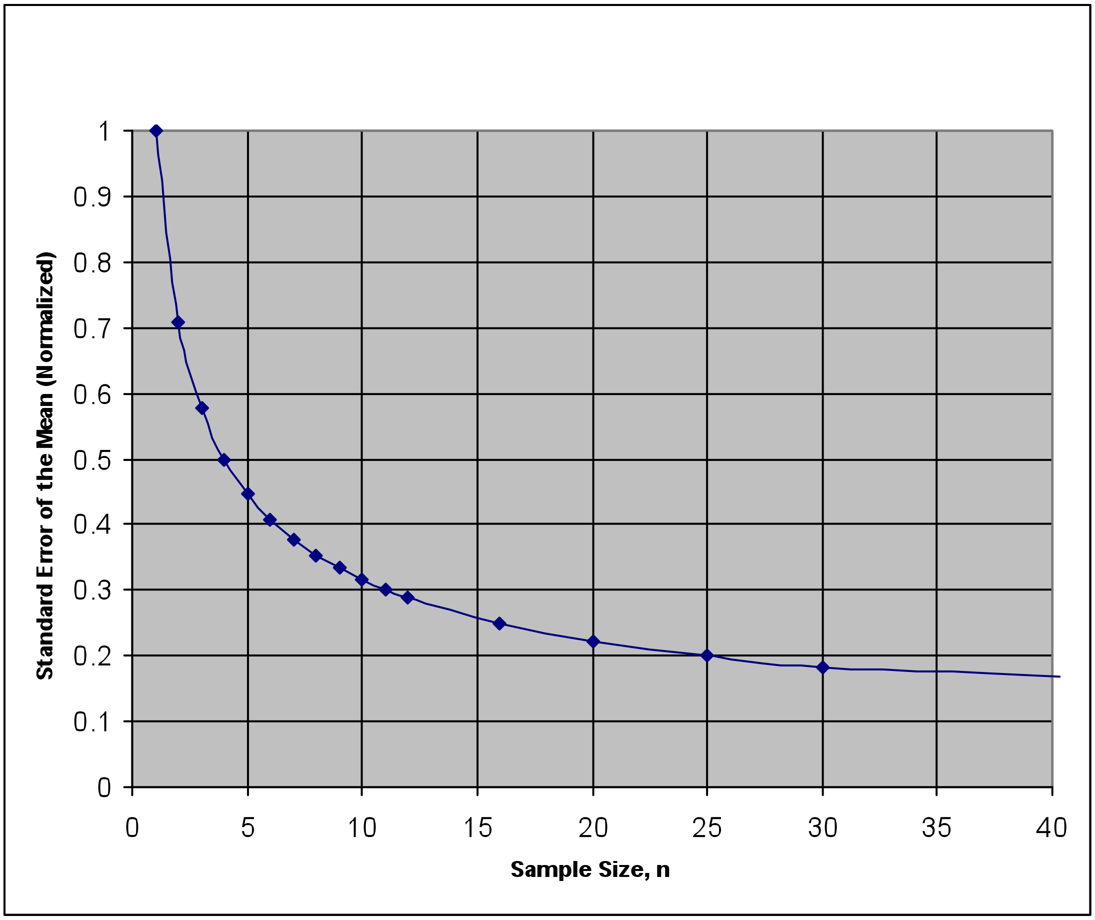


FIGURE 1 Relationship between standard error of the mean and sample size.

2. A second way was to use equations. Two equations that have been used to determine required sample sizes in highway construction are:

$$n = (Z_{\alpha} \sigma / e)^2 \tag{1}$$

and

$$n = \left[\frac{(Z_{\alpha} + Z_{\beta})\sigma}{e} \right]^2 \tag{2}$$

where
Z_{α} is the standard normal distribution value for the required seller's risk;
Z_{β} is the standard normal distribution value for the required buyer's risk;
σ is an advance estimate of the population standard deviation; and
e is the allowable difference (i.e., tolerable error) between the results of sampling and the true value of the population mean.

In order to use equations of this type, one must input the desired risk(s) and the tolerable error, and this is done mainly through engineering intuition rather than through some optimization procedure. Highway construction QA practitioners who use such equations typically conclude that n should be 15 or more.

3. A third way to determine sample size was one based on practical considerations (primarily time and personnel constraints). From a practical standpoint, a reasonable acceptance sample size is the number of tests the technician(s) can perform in one day (assuming the lot represents one-day's production). This guidance typically results in $n = 5$ for many AQCs.

Taken together, the above three approaches led those responsible for initial QA specification development, and those who followed, to conclude that n should be at least 5, more if practical, and that more is desirable. However, these conclusions are open to challenge. A major shortcoming of the reasoning behind the conclusions is that none of the above approaches adequately addresses the *optimum n*, i.e., the most cost-effective n. In addition, both the Figure 1 method and the equation method were meant for (a) accept/reject acceptance plans rather than pay adjustment acceptance plans, (b) single acceptance plans (one AQC) rather than acceptance systems (two or more AQCs), and (c) the average as the measure of quality rather than the PWL measure or any other measure of quality.

Gharaibeh's research, on the other hand, addressed the optimum sample size n for pay adjustment acceptance plans that are part of a multi-characteristic (i.e., multi-AQC) system based on either the average or the PWL quality measure. Gharaibeh defined optimum sample size as the n that minimizes total acceptance-related costs to the state highway agency (or to the contractor in the case of an acceptance system that belongs to the contractor). Primarily due to budget limitations and challenges associated with quantifying costs, Gharaibeh's objective was simply to calculate a rough estimate of optimum n. As Gharaibeh's work progressed, it became obvious that optimum n with respect to cost to the highway agency is quite small, so small it was best to form several likely-conservative assumptions needed in the absence of specific data (i.e., assumptions that would have the effect of increasing the calculated optimum n). Yet, even with those conservative assumptions, Gharaibeh's optimum n still turned out to be relatively small.

Because of the important implementation implications of Gharaibeh's finding, more research was warranted. Follow-up research was thus undertaken to more closely examine Gharaibeh's economic decision analysis model and the assumptions it called for, in an attempt to either confirm or refute his finding. This paper summarizes that research (*3*).

THE OPTIMIZATION MODEL

Economic decision theory is based on risks (probabilities) and their cost consequences. The model used by Gharaibeh is a classic economic decision theory optimization model, examples of which can be found in a number of publications (*4*, *5*, *6*). Figure 2 illustrates the general model, which identifies the optimum sample size as the n that minimizes expected total cost.

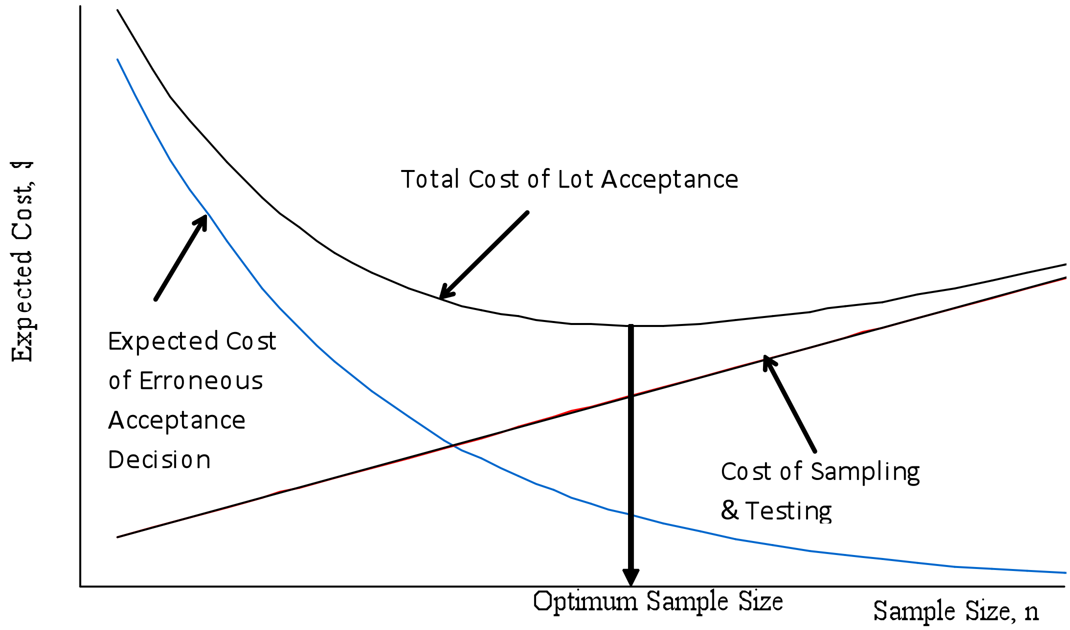


FIGURE 2 Graphical depiction of the sample size optimization concept.

From this model, one can see that the use of more costly tests would increase the slope of the "cost of sampling & testing" straight line and thereby change the "total cost of lot acceptance" curve such that its minimum point would move to the left resulting in a smaller optimum *n*. Similarly, if the cost consequences of erroneous decisions were greater, the "expected cost of erroneous acceptance decision" curve would be higher causing the minimum point on the "total cost of lot acceptance" curve to move to the right resulting in a larger optimum *n*. Both of these make intuitive sense—the more costly the test, the smaller the optimum *n*; and the greater the cost consequences, the larger the optimum *n*.

It is also important to understand that while an agency can apply the general model to either a single AQC (a plan) or to multiple AQCs (a system of plans), true optimization can occur only when the agency applies it to the whole system rather than just to a component of the system. Most, if not all, highway agencies have multi-AQC systems. The risks associated with lot acceptance of multi-AQCs are different from those of a single AQC. The more AQCs, the greater the contractor's risk (α) of having good material rejected and the smaller the agency's risk (β) of accepting poor material.

Thus, in developing an acceptance system and determining the required sample sizes to correspond to the desirable level of risk, the agency should concern itself with multi-AQC system risks. For a given set of risks, the required sample size to obtain an acceptable estimate of lot quality may be n = x for each of three AQCs in a system and, say n = 3x, for a one-AQC plan. In other words, an analysis that correctly considers the system yields a smaller required sample size for any one AQC than an analysis that is limited to a system component.

DISCUSSION OF ASSUMPTIONS

In optimizing multi-AQC acceptance plan system sample sizes with respect to cost, the objective is to determine each AQC's sample size such that the total cost associated with the agency use of its acceptance system is minimized. As can be seen in Figure 2, Gharaibeh's model has two cost elements: (1) the cost of sampling and testing, and (2) the expected cost associated with making erroneous acceptance decisions. Cost-related assumptions are necessary to graph the first element; and cost-, quality-, and risk-related assumptions are necessary to graph the second element, as the second element involves α- and β-type risks (probabilities).

An important third cost element is missing from the Gharaibeh model and was considered in the follow-up research. That third cost element is the cost stemming from construction contractors' reaction to an acceptance system that requires sample sizes different from those currently in use; in other words, is construction quality (therefore highway life cycle cost) affected if sample sizes are increased/decreased? This third cost element will be discussed later.

To establish realistic assumptions needed to determine optimum *n* in the follow-up research, the researchers used the Colorado Department of Transportation's (CDOT) construction-quality database. The CDOT database was chosen because it was found to have most of the type of data needed for the analyses. Use of the CDOT database also provided additional benefits. The database had already generated reports with data organized in such a way as to facilitate the overall optimization effort. Also, it contained a wealth of additional data to allow further related research into the effect of sample size *n* on construction quality.

A discussion follows of the critical assumptions that were made in Gharaibeh's determination of optimum *n*, the validity checks that were performed in the follow-up research, and the effect on optimum *n* of new assumptions resulting from the follow-up research.

Anticipated Quality

An assumption underlying Gharaibeh's optimization is that the agency is using/will use its acceptance system to make acceptance decisions on all incoming lots, not just on a lot of a given quality level (not on a typical lot, for example). The anticipated quality distribution of incoming lots is thus an important variable in determining optimum *n*. If incoming lots are anticipated to be generally high quality (i.e., good means, low standard deviations), less testing is needed. This is because the joint probability of a contractor delivering and the agency accepting a poor-quality lot is smaller when most incoming lots are of high quality; in Figure 2 model, the expected cost of erroneous acceptance decisions is smaller for high-quality lots.

Gharaibeh used three different anticipated lot-quality categories in his calculations: poor, regular, and good. Each category had a different distribution of AQC lot-quality levels. The three assumed distributions of lot-quality levels were such that the "poor" category had 0 percent of delivered lots that were better than the acceptable quality level (AQL), the "regular" category had 5 percent, and the "good" category had 15 percent. With these lot-quality assumptions, Gharaibeh found optimum *n* to be very small (3 or less) for all but a few cases under the "regular" and "poor" quality categories.

In the follow-up study, a check was made to determine which if any of Gharaibeh's three categories came closest to quality levels achieved in actual practice. From analysis of CDOT database reports, it was found that the AQC lot-quality distributions contractors were delivering to CDOT were much higher than Gharaibeh assumed even under his "good" quality category. The CDOT distributions were found to be such that at least 50 percent of lots were estimated to be better than AQL. Based on the CDOT data and the authors' experience with databases of other state highway agencies, there is sufficient reason to believe that most states have lot-quality distributions that are much better than the distribution Gharaibeh assumed for his "good" category.

Pay Adjustments

The theory behind acceptance plan development that was initially presented to the highway community in the 60's (*7*, *8*) dealt strictly with accept/reject acceptance plans. In the intervening years, the theory had to be expanded as state highway agencies adopted first pay decreases (penalties) then also pay increases (bonuses) associated with different levels of estimated quality. Although the vast majority of acceptance plan systems now contain pay adjustment provisions, there are still gaps in our understanding of pay adjustment development and function.

Gharaibeh assumed that the cost consequences of erroneous pay decisions were such that, in the long run, the positive cost consequences incurred when the agency underestimates quality and pays less than it should cancel out the negative cost consequences incurred when it overestimates quality and pays more than it should; in other words, Gharaibeh essentially treated pay-adjustment acceptance plans as if they were accept/reject acceptance plans.

To calculate the cost of erroneous pay decisions in the follow-up research, it was assumed that the desired pay factor for a lot having any combination of AQC PWLs is, by definition, whatever the agency wants it to be for that particular AQC PWL combination. The CDOT pay equations were thus used to determine the desired multi-AQC pay. The SpecRisk software program (*9*) was used to determine the contractor's multi-AQC expected pay. The difference between the expected pay and the desired pay is the cost of erroneous pay decisions.

The follow-up research showed Gharaibeh's assumption of zero cost of erroneous pay decisions was incorrect. As pay adjustment acceptance plans were not designed to make erroneous pay decision cost consequences cancel each other, it would be coincidental for that to happen.

The follow-up research showed that the net consequences of erroneous pay decisions not only could be considerable, but that they tended to favor the highway agency, i.e., to underpay contractors. When the agency uses an acceptance plan system that in the long run underpays contractors, the agency's expected cost of erroneous acceptance decisions (hence, the agency's total cost of lot acceptance) is decreased.

Figure 3, developed from CDOT data, can be used as an example to illustrate such a situation. The only difference between Figure 3 and Figure 2 is that the expected costs of erroneous acceptance decisions were found to be negative when developing Figure 3, making the total cost of lot acceptance negative for all *n* less than 12. The smaller the *n*, the greater the probability of an erroneous acceptance decision, thus the more negative the expected cost, i.e., the more the contractor is underpaid. These negative costs bring the optimization model's "total cost of lot acceptance" curve below the "cost of sampling and testing" curve. With respect to optimum *n* in Figure 3, it is *n* = 3, which is the lowest possible *n* for the PWL quality measure.

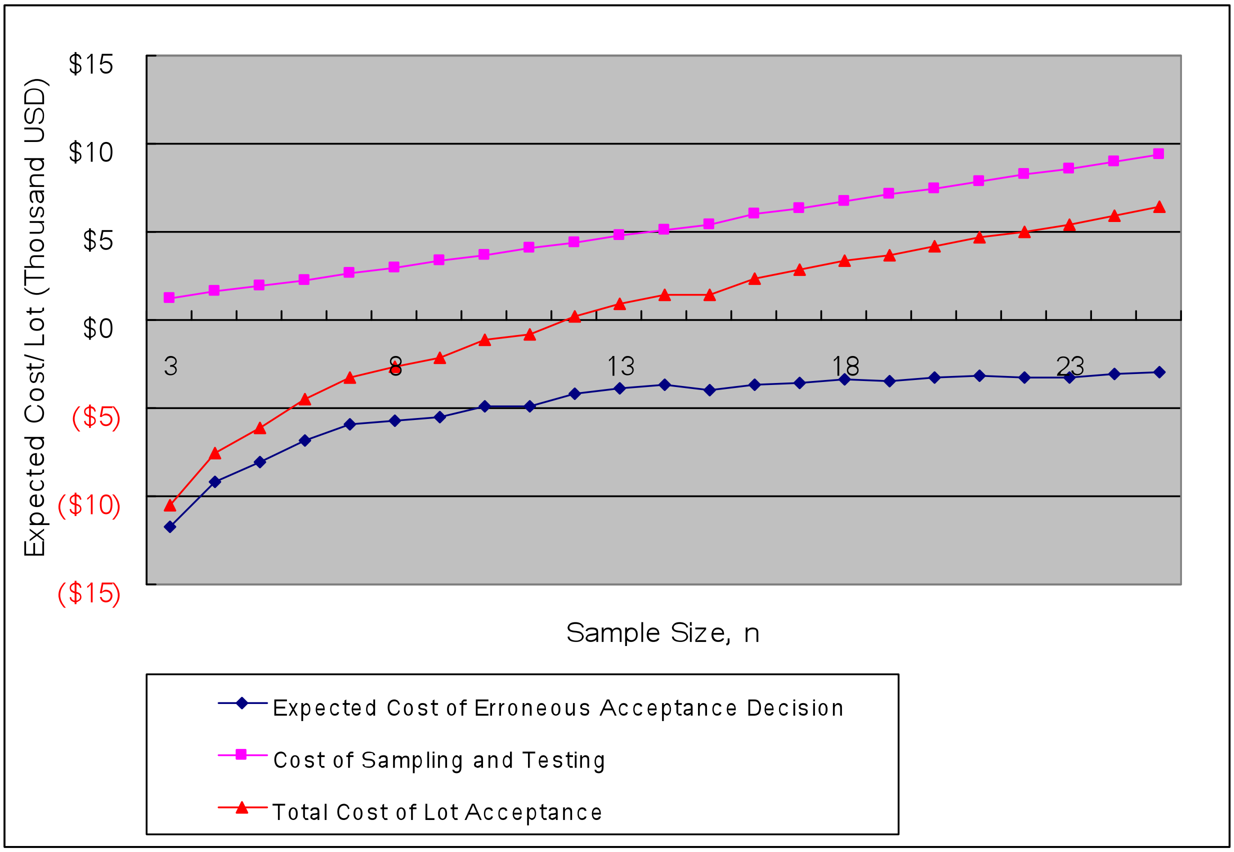


FIGURE 3 Application of optimization model with negative agency acceptance costs.

Two variables come into play for the situation depicted in Figure 3 to occur—the lot-quality distribution and the pay adjustment provisions. Figure 4 can provide a better understanding. It compares the composite pay equation with its expected pay curve for a two-AQC acceptance system (with the X-axis identifying the same PWL for both AQCs). Note that the composite pay equation has both a minimum acceptable estimated quality level (45 PWL) and a maximum pay factor cap (102.50 percent). When the lot-quality distribution is such that delivered lots rarely fall below 45 PWL, the portion of the graph to the right of 45 PWL carries a much greater weight. One can thus see that under such conditions, unless one or two specific contractors are consistently delivering the few instances of below 45 PWL lots, contractors as a whole are in the long run underpaid in comparison to agency-desired pay.

Of the two variables, the lot-quality distribution appears to be the most influential. Application of the optimization model shows that the better the lot-quality distribution, the greater the likelihood of negative costs and associated long-run underpayment to the contractors. Further, the lot-quality levels need not be all that high for the long-run underpayment to occur. It can occur with Gharaibeh's "regular" and "good" historical quality distribution, whether or not the pay provisions include a maximum pay cap and/or a minimum acceptable estimated quality level (below which lots may be rejected), although use of the latter decreases the long-run underpayment.

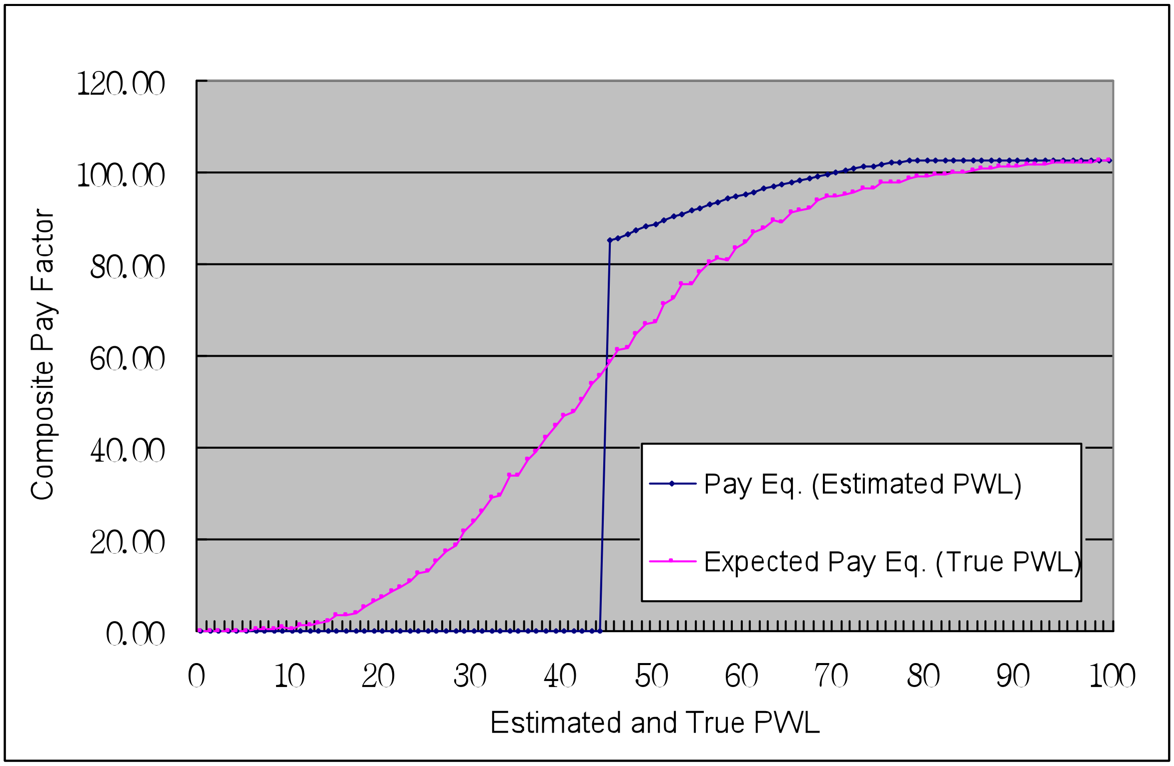


FIGURE 4 Comparison of a composite pay equation and its expected pay curve.

It is also important to note that sample size is not causing the underpayment. Sample size simply affects the risks associated with contractor being assessed the correct payment. For high-quality lots, the risk of an occasional incorrect, low-quality estimate is greater with small sample sizes, thus resulting in the pay equation bias that leads to a greater expected underpayment. The underpayment should not be an issue, provided the agency has developed, is satisfied with, and has made available to contractors, the expected pay equation.

Correlation

Gharaibeh's optimization assumed AQCs are independent or weakly dependent. The assumption seemed logical considering highway agencies are generally discouraged from using correlated AQCs (*10*). Even when agencies do use correlated AQCs, contractors tend to pay attention to individual AQC quality rather than to combined measures of quality; if one AQC is poor-quality, it does not necessarily indicate other AQCs are also poor-quality (*2*).

In the follow-up research, the effect of correlated AQCs on optimum *n* was investigated. It must be understood that, for purposes of the optimization, the issue is not whether the AQC test results are correlated but whether the AQC lot quality measures (e.g., lot PWLs) are correlated. The investigation thus attempted to determine, using CDOT data, the degree of correlation that might be expected between AQC lot PWLs, also the effect of AQC lot PWL correlation on optimum *n*.

Figure 5 is a plot of density lot PWLs versus corresponding asphalt content lot PWLs. Each data point represents a pair of PWLs that came from the same physical lot. Density and asphalt content were sampled at different frequencies, but the average sample size (average n = 24 for density and average n = 13 for asphalt content) is large enough that the PWL estimates can be considered to be adequate estimates of the population PWLs.

Figure 5 is typical of the plots one obtains with the various combinations of CDOT AQC lot PWLs. Data points are clustered at the upper right, indicating very high lot quality; and they also tend toward either axis, indicating a lack of correlation. Because conventional correlation techniques should not be used to determine the degree of correlation between non-normally distributed variables whose values contain testing error and are "top-heavy," a procedure to calculate a new measure of correlation, the "relative correlation ratio," was developed in the follow-up research specifically for use in determining the degree of correlation between AQC lot PWLs (*3*). The procedure compares the plot from actual PWL data (e.g., Figure 5) against a plot using the same PWL data but where the PWL pairs are assigned randomly, thus representing independence (i.e., no correlation).

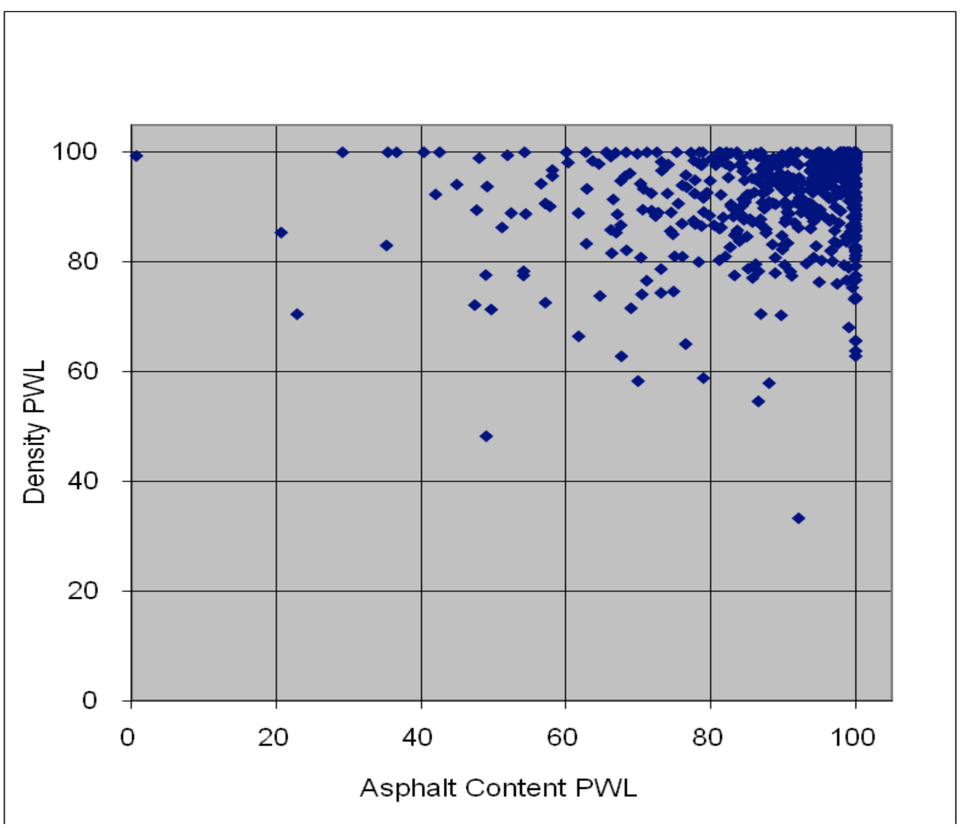


FIGURE 5 Plot of Density and Asphalt Content Lot PWL Pairs.

Overall, the degree of correlation was found to have a negligible effect on the total optimum *n*. For the degree of correlation in Figure 5 (where the plot exhibited mild correlation using the relative correlation ratio), the optimization procedure arrived at the same optimum $n = 3$ for each AQC (total $n = 6$) as that arrived under an assumption of independence. Even if the density and asphalt content lot PWLs were strongly correlated, the optimization procedure still arrives at an optimum $n = 3$ for each AQC.

However, according to the optimization procedure (and logic), if the total optimum *n* for two perfectly correlated AQCs is 6, the lowest cost to the agency occurs when $n = 6$ for the AQC having the lower sampling and testing and $n = 0$ for the AQC having the higher sampling and testing cost. Thus, unless the agency "needs" to use both AQCs, as it might when each AQC controls different key distresses, the recommended course of action to achieve optimization is for the agency to drop the higher cost AQC from its acceptance system. The optimization procedure thus provides an economic argument against the use of highly correlated AQCs (for example, in certain cases, laboratory air voids and field density).

Cost of Sampling and Testing

The cost of sampling and testing is difficult to model because there are many different possible cost scenarios: a technician could perform two or more AQC tests simultaneously; a technician could perform another function while waiting for test results; one sample could yield several AQC test results; multiple technicians could be employed and be more (or less) efficient than a single technician; etc. Gharaibeh assumed a linear relationship between the cost of sampling and testing and sample size *n*. If a single assumption has to be made about the cost of sampling and testing element in a generalized optimization model, it would be hard to argue that linearity should not be the assumption.

The assumption of linearity tends to "average" all the possible cost scenarios that exist within the state and/or within the acceptance plan system. However, if an agency believes the cost of sampling and testing element to be other than linear for its optimization purposes, the agency can substitute its own specific model in performing the necessary calculations.

In any case, agencies should also consider stepped functions. A stepped cost of sampling and testing function is appropriate in a situation where the cost of performing additional tests is viewed as negligible up to a certain *n*, at which point it suddenly increases. One such example is the cost of nuclear density testing. Under certain situations, the cost of nuclear density testing could stay about the same whether say, 3 tests or 10 tests, are performed in one day or by one technician. The cost could then noticeably increase if $n = 11$ tests required two technicians or extended the time on the job for one technician from say, 1 day to 2 days. With a stepped function, the best optimization solution is for the agency to accept $n = 10$ as the optimum in this example rather than the calculated $n = 3$.

Another assumption made by Gharaibeh dealt with the unit costs of sampling and testing. As indicated earlier, the higher the unit costs, the larger the optimum *n*. The unit costs Gharaibeh used represent national averages and are deemed to be sufficiently conservative; therefore, no further study was performed. State highway agencies that suspect their unit costs are higher can easily input their own costs.

Post-Construction Costs

Gharaibeh's optimization modeling assumed the agency's expected cost due to erroneous acceptance decisions is based solely on bid price and not any other costs such as user costs or maintenance and rehabilitation costs. This assumption may be valid for accept/reject acceptance plan systems for which there is a clear line between rejectable and acceptable lots and no need to distinguish among various levels of acceptable quality. This assumption, however, has no place in pay adjustment acceptance plan systems.

The follow-up optimization modeling by its nature distinguishes among various levels of acceptable quality (and also among various levels of unacceptable quality). It indirectly considers user costs and maintenance and rehabilitation costs. These post-construction costs relate to the expected underpayment/overpayment to contractors as a result of erroneous pay decisions.

As stated earlier, the follow-up research found that the high-quality (above AQL) lots typically delivered to state highway agencies along with the caps placed on the maximum pay factor are responsible for the long-run underpayment to the contractors. Including user costs and maintenance and rehabilitation costs in the modeling significantly raises the performance-related pay increase the contractor "deserves" to have for above AQL lots. PaveSpec performance-related specifications (PRS) software shows that the inclusion of post-construction costs, even when using only 5 percent of the theoretical user cost, can result in very high "deserved" pay factors (*11*). That is partly why 100 percent of user costs is not used in PRS development, and why a cap is placed on the maximum PRS pay factor. Both have the effect of underpaying the contractors, i.e., of creating negative costs to the agency that in turn decrease the calculated optimum *n* (below that optimum *n* calculated without post-construction costs).

Cost of Contractor Reaction

The follow-up research considered a third cost element separate from the two elements identified in Figure 2. It addressed the impact on lot quality and cost if an agency switches to a smaller *n* than the one it currently uses.

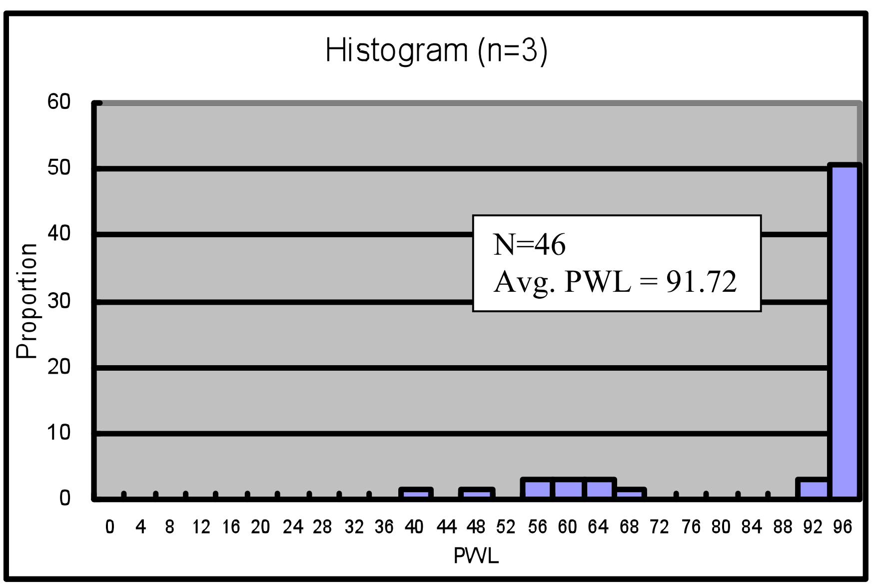
Preliminary indications are that with proper precautions there should be little if any difference in delivered (and accepted) quality. Contractors have already delivered many lots to various state highway agencies knowing that the acceptance sample size will be as small as $n = 3$ or 4.

CDOT's database was used to make a comparison of hot mix asphalt quality delivered when the contractor knows *n* will be small. The CDOT data base contains lots with *n* ranging from 3 to well over 100. The comparison consisted of grouping lots by sample size and comparing PWL estimates. To allow a fair comparison, those lots that were meant to have higher *n* but had been prematurely discontinued due to unscheduled mix design changes were eliminated from the data.

An example comparison, using the n = 3, 9, and 17-20 groups, is shown in Figure 6. As one would expect, the spread of PWL estimates decreases as *n* increases; thus, the smaller *n* distributions contain not only more low-quality estimates but also more high-quality estimates. This does not mean that the true quality is different, only that the quality *estimates* have different distributions. The average PWL of the distribution is the best measure of true quality in this case, and the three average PWLs are about the same—91.72, 90.23, and 91.87 for n =3, 9, and 17-20 respectively. The nonparametric Mann-Whitney U test shows no difference among the three population means at $\alpha = .05$ significance level.

An argument can even be made that, for some AQCs, quality increases with decreased *n*. A frequently cited example involves concrete compressive strength under an acceptance plan that uses the average as the quality measure. With smaller *n*, a contractor would need to target a higher average compressive strength in order to meet the minimum average estimated strength requirement with the same probability as for larger *n*. This argument holds not only for compressive strength but also other one-sided AQCs such as thickness, density and smoothness; and other quality measures such as PWL.

In Figure 6, it is the spread of the distributions (the error associated with the estimated PWLs) that would motivate a contractor to increase quality for either one-sided or two-sided AQCs. For two-sided AQCs, rather than increasing (improving) the lot average, the contractor would want to decrease variability. Either way, the contractor's costs theoretically would increase, but the post construction costs would probably decrease.



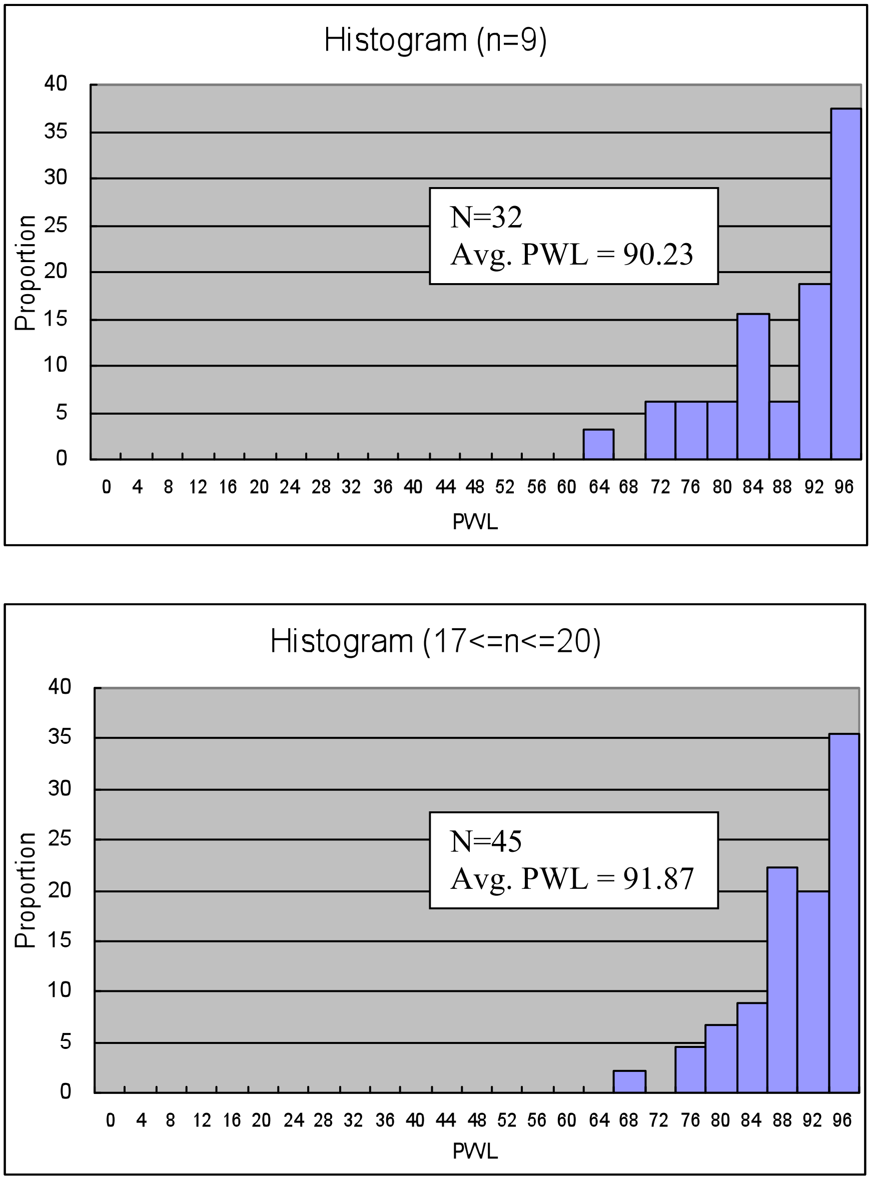


FIGURE 6 Distribution of CDOT asphalt content PWL estimates for lots with n=3, 9, and 17-20.

SUMMARY AND CONCLUSIONS

The follow-up research concluded that Gharaibeh's assumptions were indeed conservative. With modified assumptions, based where possible on actual data, the optimum acceptance sample size *n* is even smaller than that calculated by Gharaibeh. However, as Gharaibeh pointed out, *n* cannot be less than 3 for PWL acceptance plans as PWL estimates cannot be made with less than 3 test results.

For acceptance plan systems related to pavements, optimum *n* is primarily a function of incoming lot quality—the higher the quality the lower the *n*. To illustrate with an extreme example, if an agency was certain that incoming lots will be of such high quality that none should be rejected, there would be no need for the agency to do any testing under an accept/reject acceptance system. The term "acceptance system" however, has become somewhat of a misnomer. For many highway agencies, it is more a "pay adjustment system," i.e., more for the purpose of making pay adjustment decisions rather than accept/reject decisions. The follow-up research examined pay adjustment acceptance systems and concluded for them too, optimum *n* is primarily a function of incoming lot quality. When incoming lot quality is anticipated to be high, the expected cost associated with erroneous pay decisions is frequently negative (i.e., contractors are underpaid in the long run), especially in situations where the agency has no minimum acceptable estimated quality level provision (or it has a provision, but with a low minimum).

For acceptance plan systems related to other than pavements (e.g., bridge decks), it should be noted that optimum *n* may also be influenced considerably by the cost associated with erroneous accept/reject decisions. If the consequences of erroneous accept/reject decisions are catastrophic and could result in loss of lives, the optimum *n* could be much higher than 3. Both Gharaibeh's research and the follow-up research reported here considered only pavement acceptance plan systems.

Optimum *n* is of course also a function of the AQCs being measured—what are they? how many are there? are they correlated? and what is the cost to sample and test them? The number of AQCs is important as they work as a unit to estimate lot quality which then determines the composite pay factor. Because the expected cost of erroneous pay decisions is spread out among the AQCs, each AQC's contribution to the expected cost decreases as the number of AQCs increases, resulting in a smaller optimum *n* for each AQC. The follow-up research investigated systems with only two AQCs, and already optimum *n* was below 3. Correlated AQCs were not found to change the optimum $n = 3$ conclusion for PWL, provided the agency needed the correlated AQCs within the acceptance system in order to control different key distresses. The unit costs of sampling and testing derived by Gharaibeh were deemed to be conservative and were used in the follow-up research as well.

The follow-up research concluded statewide lot quality (and therefore lot performance) is not likely to suffer if agencies that switch to smaller acceptance sample sizes take proper precautions. Some recommendations are provided below in the recommendations section. There is also reason to believe quality might actually increase for some one-sided AQCs, as statistical-risk-aware contractors tend to raise target quality to account for the increased variability of estimates from small sample sizes. In such cases, to keep from having a corresponding increase in the cost of lots, agencies that expect contractors to raise their quality levels might consider simply lowering the specified quality level. However, it is doubtful that any increase in costs due to such increases in quality would change the optimum *n*. For agencies that believe quality levels will decrease, the simple solution would be to raise the specified quality level.

RECOMMENDATIONS

It is strongly recommended agencies do their own optimum *n* determinations. In doing so, they will gain a better understanding of their acceptance plan systems and associated costs, and just as important, of the various underlying assumptions and their effect on optimum *n*. Having this understanding, they will be in a better position to draw their own conclusions from the performed economic decision analysis, which identifies the best decision in the long run and not necessarily the best decision with respect to a specific project or contractor or submitted quality level. The understanding will also provide the agencies greater confidence in applying economic decision analysis principles to minimize expected costs and optimize statistical risks.

Once an agency has followed the optimization procedure and determined optimum *n* for its AQCs, the agency will have simultaneously identified the optimum buyer's and seller's risks (as these statistical risks are a function of *n*). It is also possible for the agency to determine the theoretical optimum lot size, as was done by Gharaibeh et al (*2*), since lot size affects the expected cost of erroneous acceptance decisions (an element of the optimization model).

Assuming the agency is considering switching to a lower *n* based on its optimization, the following recommendations are offered:

- In going to less acceptance testing, the agency should consider (a) placing more emphasis on the contractors' quality control programs, (b) placing more emphasis on inspection to identify isolated instances of poor quality, and/or (c) replacing programs that use contractor test results for acceptance purposes with programs that use only the agency's test results.
- The agency should include a strong retest provision in each acceptance plan to further test lots that yield borderline test results.
- If the agency can identify specific contractor(s) with a record of having submitted low quality levels in the past, the agency should consider taking appropriate action. The agency has many options, one of which is the use of an acceptance plan system with same low *n* but greater pay reductions for the specific contractor(s).
- The agency should monitor how the lower *n* is working statewide with respect to quality and cost. Here too, the agency has many options that allow it to control overall quality and cost; simply increasing/decreasing specified quality is one such option.

ACKNOWLEDGMENT

The authors thank Eric Chavez of the Colorado Department of Transportation for providing the data in a form suitable for performing the analyses in this study.

REFERENCES

1. Von Quintus, H.L., Rauhut, J.B., Kennedy, T.W., and Jordahl, P.R., "Cost Effectiveness of Sampling and Testing Programs," Federal Highway Administration, Publication No. FHWA/RD-85/030, McLean, VA, 1985.
2. Gharaibeh, N.G., Garber, S.I., and Liu, L., "Determining Optimum Sample Sizes for Percent Within Limits Specifications," Transportation Research Board, Transportation Research Record 2151, Washington, DC, 2010, pp. 79-83.

3. Cho, D., "Determining Optimum Sample Sizes with Respect to Cost for Multiple Acceptance Quality Characteristics," PhD dissertation, University of Florida, 2010.
4. South, J.B., "Selecting an Acceptance Sampling Plan That Minimizes Expected Error and Sampling Cost," Pittsburg State University, Quality Progress, Vol. 15, No. 10, pp. 18-22, 1982.
5. Lapin, L. *Statistics—Meaning and Method*. Harcourt Brace Jovanovich, Inc. New York, NY, 1975.
6. Fabrycky, W.J. and Thuesen, G.J., *Economic Decision Analysis*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1974.
7. "Military Standard 414, Sampling Procedures and Tables for Inspection by Variables for Percent Defective," Superintendent of Documents, Government Printing Office, Washington, DC, 1957.
8. Miller-Warden Associates, "Development of Guidelines for Practical and Realistic Construction Specifications," Highway Research Board, National Cooperative Highway Research Program Report 17, 1965.
9. National Highway Institute, "SpecRisk Quality Assurance Specification Development and Validation Course," Web-based course number FHWA-NHI-134070, www.nhi.fhwa.dot.gov.
10. Burati, J.L., "Risks with Multiple Pay Factor Acceptance Plans," Transportation Research Board, Transportation Research Record No. 1907, Washington, DC, 2005, pp. 37-42.
11. Hoerner, T.E. and Darter, M.I., "Improved Prediction Models for PCC Performance-Related Specifications, Volume II," Federal Highway Administration, Publication No. FHWA-RD-00-131, McLean, VA, 2000.