FICO® Score Validation Guide

January 2017

Table of Contents

Introduction

With the proliferation of credit scoring solutions available in the marketplace, many financial institutions can benefit from predictive scores but may be limited by not fully understanding how to evaluate and decide on an appropriate solution. In addition, lenders who need this kind of information are busier than ever.

This guide was written by FICO to provide your organization with "just-in-time" information on how to evaluate the industry dominant FICO® Score. It starts by first discussing organizational objectives and then moves on to the more tactical aspects of conducting a validation to ensure that you are evaluating what really counts for your business environment. The information provided in this guide is not intended to be technically exhaustive, but rather to help you design a more effective FICO® Score validation that provides meaningful results on the potential future impact of the FICO® Score for your various portfolios.

This guide answers key questions such as:

- How do I best structure a FICO® Score validation?
- How do I interpret validation results?
- How do I ensure the FICO® Score will work in the strategies I implement?
- How can I judge whether the FICO® Score demonstrates bottom-line benefits for my business?
- How do I compare the effectiveness of two different scores in my business environment?

Generally, a FICO® Score validation consists of comparing the score's ability to rank order the actual performance of accounts. A low FICO® Score should indicate a riskier individual whereas a higher scoring individual should demonstrate less risk. Various reports are then analyzed to identify the effectiveness of the FICO® Score on your portfolio as well as to determine initial score cut-off strategies. The guide covers the key steps lenders need to take in the evaluation process, including:

1. Determining organization goals and an appropriate validation design
2. Preparing the validation data sample
3. Requesting the FICO® Score from your Credit Reporting Agency (CRA)
4. Conducting the validation and analyzing the reports
5. Determining the financial impact of the FICO® Score on your portfolio

For your convenience, at the end of the guide, we include an easy-to-use checklist that summarizes the various considerations when designing a FICO® Score validation and examining results. We have also included a glossary with definitions of terms commonly used during validations.
For further information regarding the technical development of the FICO® Score, please reference the FICO® Score User Guide.

Ensuring Validation Objectives Support Management Goals

Before designing a FICO® Score validation, you must first identify and prioritize the key management objectives for your organization. For example, are you trying to maximize profit, keep charge-off rates below a certain level, increase revenue or increase your client base in a particular segment?

Your management goals will influence your validation design. To illustrate this, the following table lists several examples of validation designs based on possible management goals.

Management Goal	Sample Validation Design
Lower charge-offs on specific portfolios	Compare FICO® Score versus current score on overall populations as well as by population segment on existing accounts (e.g., those with high-revenue potential, or by portfolio or product type)
Forecast bad rates for the next six months and year for financial projections	Calibrate performance on more than one time window (e.g., for six months and a year)
Cost-justify the use of a score	Use statistical measures of FICO® Score effectiveness and translate results into bottom-line dollar measures (see **Evaluating the Bottom Line Impact**)

Recognizing conflicting goals

Your organization might have several objectives with respect to a given portfolio, and these goals may sometimes be in conflict with one another. For example, an objective to maximize revenue may be in conflict with one to reduce charge-off rates, if the population segments with high charge-off rates contribute a high amount of revenue. In this case, you might combine these objectives by evaluating how well the FICO® Score would reduce charge-off losses while minimizing impact to revenue streams.

Identifying other validation objectives

You may have other business objectives that will also affect the validation design. For example, if one goal is legal compliance, you can construct an odds-to-score relationship to demonstrate that the FICO® Score rank-orders your population by risk (see Figure 5). If the purpose is to determine strategies for optimal usage of the FICO® Score, validations can be used to simulate alternative strategies used for implementing the score. To cost-justify an additional analytic tool, a cost-benefit analysis can determine how the dollar value of any improvements derived would compare with the cost of the FICO® Score and its implementation. A cost-benefit analysis is also important when comparing competing scores for use in your organization. Keep in mind that this test must be designed to:

- Be objective, comparing all scores on an equal basis. For example, an independent validation sample should be used such that no one score is favored (see **Validate on an independent sample**). Common scores to compare include a prior version of the FICO® Score or a custom developed score.

- Simulate the environment in which the FICO® Score will be implemented. This includes the populations on which it will be applied, the context of strategies into which it will be incorporated and the performance measure(s) most relevant to your organization.

Preparing the Validation Data Sample

Once you have determined your organization's objectives, you are ready to prepare your data for validation.

Identifying an appropriate data source

Depending on where you plan to use the FICO® Score, you will first need to identify a data source appropriate for the validation. Considerations should be made for various portfolios, product types, and uses of the FICO® Score. For example, if you plan to use the FICO® Score in your originations decision, an appropriate set of applicants will need to be identified. For account management decisions, lenders will need to choose from an existing set of accounts. Selection of these applicants or accounts will depend on several factors and may be influenced by portfolio dynamics.

For each of these consumers, the necessary identifying information needed to obtain a credit report must be provided. In addition, you will need to include the actual performance of the account, or how the consumer performed on the account over the performance window. Periods of time that introduce an unusually different profile of applicants or accounts should be avoided, such as a new marketing campaign for a new population segment. It is also recommended to avoid seasonal swings in volumes, if possible.

Performance Definition – Defining what to measure

When designing or evaluating a validation, the performance definition you establish for the study may affect which score the validation determines as "better." Therefore, you need to establish performance definitions that are aligned to your organization's business objectives. For example, in a validation that assesses how well the FICO® Score separates "good" and "bad" accounts, you will need to define good and bad in a way that makes sense for your business.

For existing accounts, performance can be determined based upon data available within your masterfiles. For your applicant population, performance can only be determined for applicants on which you have observed data – your booked accounts. For performance on rejected (e.g. not accepted) and un-cashed applicants (e.g. accepted by not taken up by consumer), you can infer performance by requesting from your CRA a performance proxy flag for that consumer based on similar performance on another tradeline. Please see **Dealing with Truncated Data** for additional information on handling rejected and un-cashed applicants.

The performance information will be used to determine how well the FICO® Score rank-orders the risk of your population. For applicants, key information to determine performance may be the approval decision (accept/reject) for each application, as well as the payment performance for all booked accounts over a period of at least 12-24 months. For existing accounts, performance may be determined by the payment performance over a period of at least 12-24 months. A suggested layout of performance data is provided in Appendix C.

Specifically for the FICO® Score, the development performance definition measures 90+ days past due or worse over 24 months from scoring. Depending on your own portfolio characteristics, you can choose to use the same performance in your validation or modify it to be more or less severe. Modifications may also be necessary if there are insufficient counts to use the performance definition of choice.

Comparing scores on the same performance definition

A score developed using a particular performance definition does not have to be validated on that same measure. In fact, you should not compare two scores built on different performance definitions on their respective performance measures. This would be an "apples-to-oranges" comparison and not analytically sound.

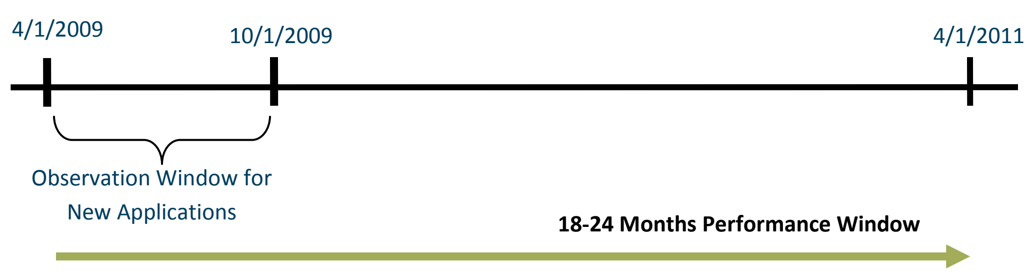
Instead, the key is to use a performance definition that is meaningful to your business' operations. When comparing multiple scores, you should evaluate them on the same performance measure—thus making it an "apples-to-apples" comparison using standards that would be useful to your organization. For instance, you might want to compare an existing risk score (with a good vs. 120+ days delinquency development definition) to the FICO® Score (with a good vs. 90+ days delinquent development definition) on your own performance definition of good vs. bankruptcy.

Performance Window – Determining how long to measure performance

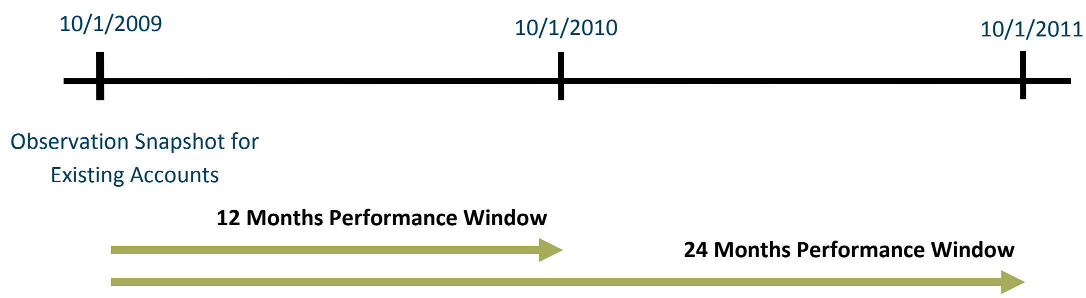
The performance window refers to the period of time over which performance will be evaluated. For example, if an account was booked in October 2009, we could evaluate their performance from November 2009 through April 2011 for an 18 month performance window.

Consider optimal length of time

The performance window used in the validation should match the practical period in which account performance will be measured. For example, the FICO® Score may be used in decisioning for booking new accounts. Since delinquencies generally peak 18–24 months after the accounts are booked, this timeframe may be the appropriate performance window.



For decisions on existing accounts, observations are usually chosen for one to two points in time. Delinquencies can then be observed 12 – 24 months from the observation date.



Sample Size - Determining how many records you need

The final count is determined by factors such as the application volume, approval rate, account volume and bad rate for each portfolio. Generally, you need to maintain a minimum number of *bads* in your sample to ensure a robust sample. Depending on the size of your applicant population or booked account portfolio, you may choose to select a subset of applicants or accounts for the analysis. The final data should include at least 500 *bad* accounts for each group that should be evaluated (e.g., you should evaluate your auto portfolio separately from your credit card portfolio). For example, if you are evaluating applicants and have an approval rate of 70% and a bad rate of 5%, you would need over 14,000 applications in the sample in order to obtain the minimum of 500 *bads*. Note that additional *bad* accounts are required if the validation will include comparing multiple scores, such as an internally developed score with the FICO® Score – with ideally at least 1,500 bad accounts for a dual-score comparison. If the applicant or account population is larger than necessary, a common practice is to sample all the *bad* accounts and randomly sample or stratify sample a portion of the *good* accounts to reduce the population to a manageable size.

Sample Window – Choosing a time period to analyze

The sampling window refers to the period of time over which applications or accounts are selected for the validation. For example, you may take all applications from April 2009 –October 2009 for the validation. For existing accounts, you may take all accounts that are on the books at a single point in time (e.g. October 2009). The length and timing of the sampling window will be influenced by the bad rate and volumes for your portfolio(s). If your bad rate or volume is low, then you may need to have a longer sample window. Seasonality and data anomalies should be also be considered when choosing a sample window.

Sample population selection is also an important component of validation design. If the sample is not chosen properly, the effectiveness of the FICO® Score might not be accurately measured.

Validating on an independent sample

The sample used for the validation should be independent of the development samples used for any of the scores being tested. For example, you may be comparing the FICO® Score against an existing score developed on your data. Often times, a score is particularly fine-tuned to its development sample, but may not hold up as well when validated on an independent sample. Thus, when comparing the FICO® Score against an existing score, you should ensure the validation sample chosen is independent of the existing score's development sample. Otherwise, the existing score may have an advantage, when compared to the FICO® Score; however, this does not guarantee it would perform better upon implementation.

Exclusions – Selecting an actionable sample population

Certain types of applications or accounts should not be used in the validation as they may bias the results or do not represent the population which will be evaluated with the FICO® Score. For originations, this generally includes any applications that are policy declines/accepts or overrides to strategy decisions. Policy declines/accepts are applicants that are declined or accepted automatically based on an established rule. For example, an applicant with a poor previous history with the bank may be an automatic decline, regardless of their current credit history. Because the FICO® Score will not be used to evaluate these cases, they should be excluded from the validation sample or clearly identifiable so that they can be removed. Similarly, in cases where lenders have policy accept rules, which automatically approve an applicant regardless of the credit history, (e.g., bank executives), the applicant should be excluded or clearly identified.

In addition, accounts with performance that does not represent actual account holder performance (e.g. lost/stolen, fraud) should be excluded. It may also not be appropriate to include accounts that are no longer "actionable", (e.g. 90-days delinquent at observation) —that is, the accounts are past the point where you can prevent further losses. It would be better to focus on the FICO® Score's ability to identify delinquency several months in advance.

For additional exclusions applied during the development of the FICO® Score, please consult the FICO® Score User Guide.

Dealing with Truncated Data

As mentioned above, you should ideally evaluate the FICO® Score on a population similar to the one on which it will be applied. However, because of decisioning or scoring processes used to acquire your customers, you may be limited to using a subset of your actual "through-the-door" population—in other words, you may have a truncated sample.

Truncation of your validation sample may be difficult to avoid, but you should nevertheless be aware of its potential effects. You may consider booking a small control group of accounts that would have been rejected by your usual decision criteria, and include these accounts in the validation sample. While a control group may have an impact on loss rates, it provides you with more accurate validation results and a better assessment of score's performance upon implementation which may have longer-term benefits. Another option is to request proxy performance from the bureau, for consumers in your "through-the-door" population for whom you do not have on your books.

Additional Data Elements

In addition to the identifying information, and performance of accounts, other data items may be relevant for the validation. These additional items may be used for comparison (e.g., previous scores), segmentation of results (e.g., portfolio, product type), or verification (e.g., account open date). A set of suggested fields follows, with an example layout provided in Appendix C.

- Existing Score(s) (e.g., an application score currently used for decisioning or a prior FICO® Score)
- Override Reason
- Portfolio Indicator
- Date Account Opened
- Date Account Closed
- Account Closed Reason
- Account status

Generating the FICO® Score

After the validation data file has been created, the information should be passed to your credit reporting agency (CRA) of choice. The CRA will then generate the FICO® Score as near to the observation date as possible based on historical archives. The date of score generation should precede the date of application, if you are evaluating an applicant population. For the cleanest score comparison, any prior FICO® scores that will be included in the evaluation should also be generated by the CRA at the same time, on exactly the same data.

A depersonalized file is then created by the CRA that appends the generated scoring output to your file containing the performance variables and other scores and/or variables to be analyzed. It's important to make sure your validation file contains all of the variables you'll need to perform the validation, as you'll not be able to match it back to your account data once it has been depersonalized.

Conducting the Validation

Validation Reports

A variety of statistical measures and reports can be produced to quantify the performance of the FICO® Score and provide you with a better understanding of its impact on your portfolio. The following section discusses commonly used reports and statistical measures that indicate the predictive value of a score, and graphical reports that demonstrate how the FICO® Score will perform. Once the validation is complete, these reports can also be used to help you integrate the scores within your strategies – in setting cut-offs and designing strategies.

Summary Statistics

In order to assess how effective the score will be in practice, it is useful to understand what different measures evaluate and how to translate this information into results that are meaningful to your organization. Each measure has its advantages and limitations.

Divergence

- Definition: Summary metric measuring the separation between distributions of two groups of accounts (e.g., "good" vs. "bad" accounts) by the FICO® Score. The better score is able to better separate the two populations, thus having a higher divergence. Divergence is calculated as follows:

$$\frac{\left(\mu_{Good} - \mu_{Bad}\right)^2}{\frac{1}{2}\left(\sigma^2_{Good} + \sigma^2_{Bad}\right)}$$

- Advantages: Measures across the entire FICO® Score distribution. Takes into account the separation of the two groups and the variances of the distributions.

- Limitations: When comparing two scores, it does not necessarily identify the better score at a particular operating threshold.

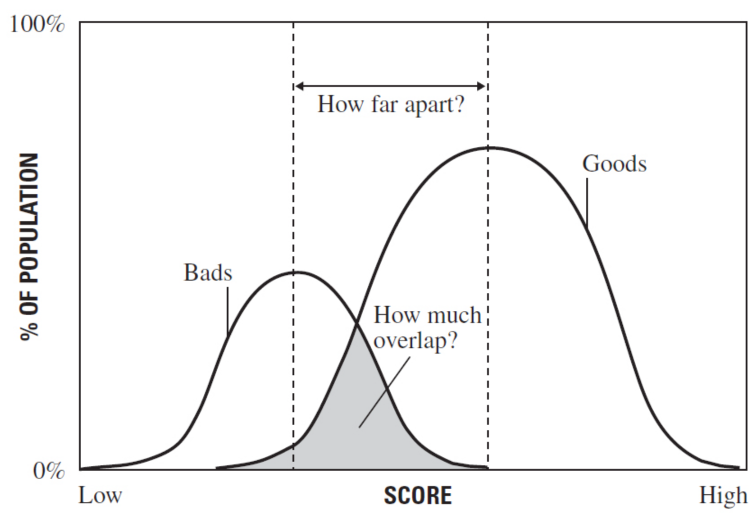


Figure 1: Divergence

Kolmogorov-Smirnov statistic (K-S)

- Definition: Measure of the maximum difference between the cumulative percentage of two groups of accounts (e.g., goods and bads) by FICO® Score.
- Advantages: Measures the point of maximum separation between cumulative distributions. Visually easy to interpret.
- Limitations: Only measures separation at one point in the distribution. The point of maximum separation may not be within your operating threshold. If there is insufficient data/insufficient sample size, the measure may be overstated / inaccurate.

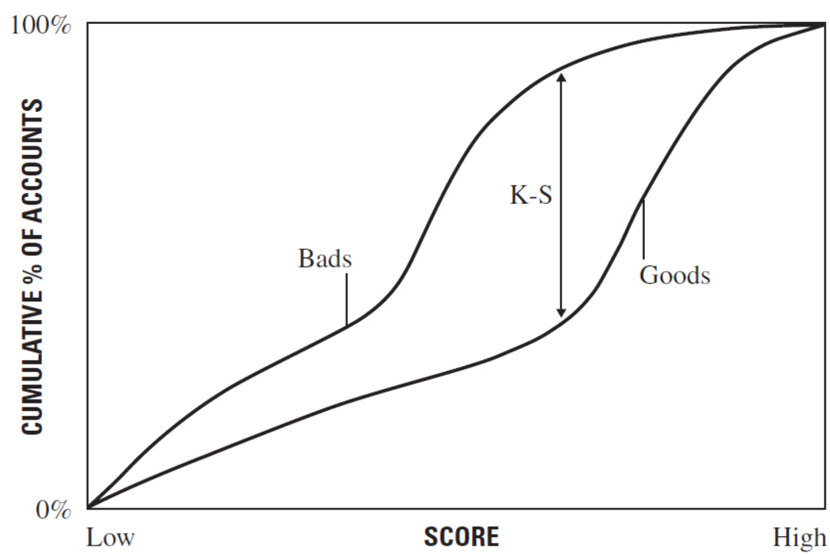


Figure 2: Kolmogorov-Smirnov (K-S) Statistic

Receiver Operating Curve (ROC)

- Definition: Summary measure of the area under a trade-off curve. The larger the area, the better the score.
- Advantages: Provides a quantitative value. Helpful when comparing trade-off curves across a population. In instances where trade-off curves cross over each other, ROC helps indicate which score is more effective overall.
- Limitations: May not indicate the better score at a particular operating threshold.

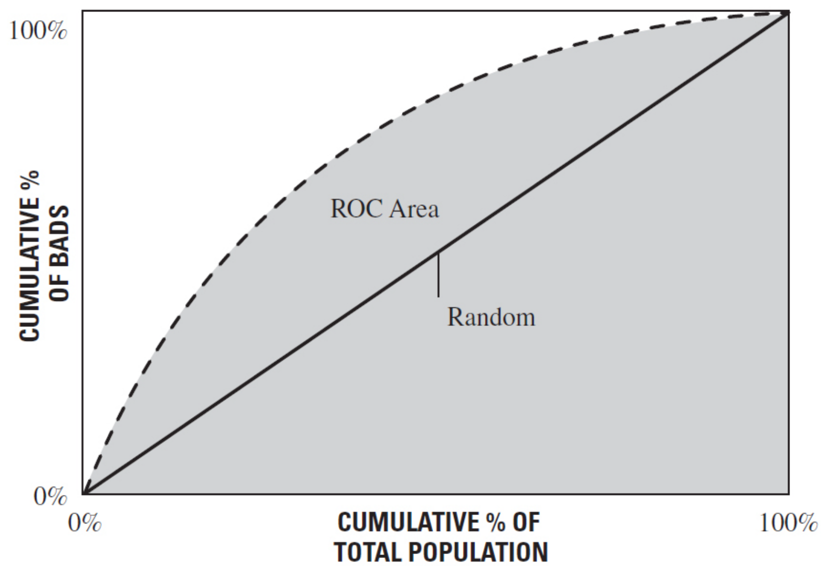


Figure 3: Receiver Operating Curve (ROC)

Trade-off Curve

The trade-off curve plots the ascending accumulation of one group of accounts vs. another group of accounts. It is also often known as Lorenz curve, good/bad trade-off curve or lift curve. This report is useful for visually comparing the FICO® Score against an existing score's effectiveness at a particular operating point or across the spectrum of the score distribution. It is independent of the score scale, allowing for equal comparison of two different scores. For example, in the graph below, at X% of the cumulative total population, the FICO® Score identifies Y% of cumulative bads, and Score B identifies Z% of cumulative bads. Thus, the FICO® Score is the more effective score as it identifies a greater percentage of bads. However, when comparing two scores with trade-off curves that overlap several times, it's often difficult to assess which score is better.

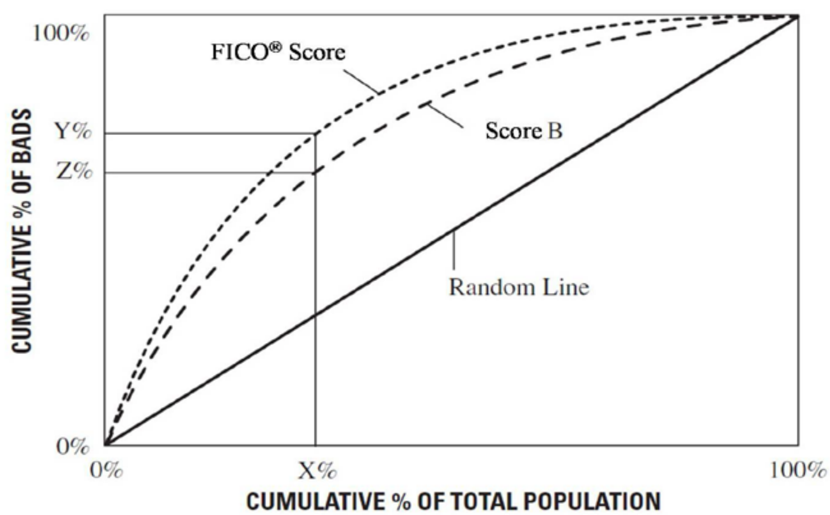


Figure 4: Trade Off Curve

Odds-to-score slope/relationship Report

An odds-to-score report plots the log (odds) at each FICO® Score point or score range. This report can be used to demonstrate the rank-ordering ability of the FICO® Score. It can be used to demonstrate legal compliance or to benchmark/ compare the FICO® Score to itself over time. However, since the scaling of different scores can be arbitrary, this report cannot be used to compare the performance of two scores.

Please note that the validation odds that you will observe on your portfolio will differ from the published developmental odds provided for the FICO® Score. FICO® Score validation charts display the performance for consumers represented in the national development sample for each CRA. They reflect the combined experience of many credit grantors in the aggregate and do not provide a precise prediction of the actual odds that a given credit grantor will experience.

Figure 5: Odds-to-score Report

Performance/Bad rate Distributions by Score Band

This report identifies the number and percentage of accounts in each score band along with the odds or bad rate observed in each score interval, based upon your chosen performance definition or for a variety of different credit performance definitions.

Bad rate distributions can be used to validate the rank-ordering relationship between the FICO® Score and the various levels of credit performance for your portfolios. Once actual portfolio performance expectations are understood by score range, actual score distributions on applicants, prospects or customers, this report can be used to select score cutoffs and design strategies.

	EVALUATING PERFORMANCE BY REPORT					
	BASE CATEGORY		NEGATIVE PERFORMANCE		%	
FICO® SCORE RANGE	# OF CONSUMER REPORTS	CUMULATIVE %	# OF CONSUMER REPORTS	CUMULATIVE %	NEGATIVE TO BASE (REPORTS)	REPORT ODDS (VALUE to 1)
<500	1,796	2.5	456	14.5	25.4	2.9
500-519	1,197	4.1	249	22.4	20.8	3.8
520-539	1,530	6.2	246	30.2	16.1	5.2
540-559	2,155	9.2	296	39.5	13.7	6.3
560-579	2,730	13.0	307	49.3	11.2	7.9
580-599	3,183	17.4	281	58.2	8.8	10.3
600-619	3,633	22.4	260	66.4	7.2	13.0
620-639	4,063	28.0	232	73.8	5.7	16.5
640-659	4,538	34.3	211	80.5	4.6	20.5
660-679	4,937	41.1	181	86.2	3.7	26.3
680-699	5,265	48.3	150	90.9	2.8	34.1
700-719	5,608	56.1	108	94.4	1.9	50.9
720-739	5,721	64.0	74	96.7	1.3	76.3
740-759	5,556	71.6	45	98.1	0.8	122.5
760-779	5,339	79.0	28	99.0	0.5	189.7
780-799	5,683	86.8	18	99.6	0.3	314.7
800+	9,541	100.0	13	100.0	0.1	732.9
TOTAL	72,472	100.0	3,155	100.0	4.4	22.0

Table 1: Example Performance/Bad Rate distribution by Score Band

Cross-tab Distribution Report

Often, the existing score that a lender may compare against will also be a prior version of a FICO® Score. When transitioning to newer FICO® versions, score distribution shifts are inevitable. Typically, only a small percentage of the consumer files receive exactly the same score. Some files score higher and some score lower. Generating a cross-tab distribution matrix will help with evaluating the potential changes to account volume shifts relative to score cutoffs. For example, a strategy to maintain risk performance may experience shifts if using the same score cutoff. However, you can analyze this report and the current distribution across the two scores to meet specific approval volume targets and help determine where to set the cutoff.

Score Movement Comparison
FICO® Score vs Prior FICO® Score*
All Industries - All Accounts

Prior FICO® Score	FICO® Score										
	300 - 499	500 - 539	540 - 579	580 - 619	620 - 659	660 - 699	700 - 739	740 - 779	780 - 819	820 - 850	Row %
300 - 499	50.14%	32.68%	14.46%	2.56%	0.14%	0.01%	0.00%	0.00%	0.00%	0.00%	5.2%
500 - 539	27.32%	32.87%	28.67%	9.99%	1.01%	0.11%	0.02%	0.01%	0.00%	0.00%	5.9%
540 - 579	18.41%	21.89%	30.56%	23.17%	5.35%	0.47%	0.10%	0.03%	0.01%	0.00%	6.6%
580 - 619	6.71%	10.47%	21.54%	34.81%	22.74%	3.24%	0.37%	0.08%	0.03%	0.01%	6.9%
620 - 659	0.67%	3.26%	6.42%	22.03%	40.41%	22.45%	4.33%	0.34%	0.07%	0.01%	8.2%
660 - 699	0.02%	0.58%	1.16%	4.94%	21.69%	47.38%	20.68%	2.48%	1.01%	0.05%	10.3%
700 - 739	0.00%	0.02%	0.08%	0.39%	3.30%	19.09%	55.92%	17.99%	2.98%	0.23%	13.7%
740 - 779	0.00%	0.00%	0.00%	0.01%	0.26%	3.31%	17.98%	51.33%	25.85%	1.25%	17.0%
780 - 819	0.00%	0.00%	0.00%	0.00%	0.01%	1.17%	4.83%	16.30%	64.26%	13.43%	25.1%
820 - 850	0.00%	0.00%	0.00%	0.00%	0.00%	0.09%	2.35%	2.88%	41.59%	53.08%	1.2%
Column %	6.0%	6.1%	6.6%	7.0%	8.0%	10.5%	14.5%	15.6%	21.5%	4.2%	100.0%

*Note: These numbers are for illustrative purposes only and do not reflect actual results.

Table 2: Example Cross-tab Distribution Report

Validating on Key Segments of your Organization

When you design validations or review results, you may want to consider the validation on key population segments in addition to the overall population. This will help you evaluate the FICO® Score's effectiveness on those subpopulations of interest to your organization. Evaluating the FICO® Score on particular segments of the overall population can lead to insights about new opportunities for profitability in these market niches, as well as the behavior of particular population segments.

Understand which populations the scores were designed to treat

While two scores may exhibit similar effectiveness when applied to a broad population, development features of one score may make it clearly superior for specific market niches. When presenting validation results, analysts need to focus on these segments, or subpopulations, if they are critical for your business. For example, if targeting the underserved market is of strategic importance for your organization, then evaluating the score on a thin or young population may provide you with greater insight on how effective that score will be at addressing a key market segment.

Evaluating the Bottom Line Impact

When comparing the FICO® Score against another score, using dollar-amount assumptions about good and bad accounts can help convert differences between scores into bottom-line benefits that are meaningful to your organization. This analysis can be particularly useful when one of the purposes of the validation is to justify the purchase of the FICO® Score or to justify the effort of implementing a new score.

Actual or estimated values for revenues and expenses, as well as information in your performance reports or summary statistics, can be translated into profitability results. For example, you can use information derived from validation data plotted on a trade-off curve to do this type of analysis (see Figure 4). Assumptions can be made about losses on bad accounts, revenues on good accounts and total fixed or variable costs. These can then be applied in simple strategies. A lender could also apply the financial figures in a more sophisticated manner—for example, varying the revenue or loss amounts by score range. However, many institutions have difficulty accessing complete financial data, particularly allocating overhead expenses. How accurate you are in estimating these figures will influence how well you can estimate the FICO® Score's contribution to your bottom-line results.

It is important to stress that when comparing scores, variations in K-S, divergence and other performance measures may seem small, but may translate into a substantial impact on your volumes and bottom line.

Evaluate implementation strategies

Be sure to examine validation results in the context of possible implementation strategies, not just on a standalone basis. In reality, the FICO® Score is likely to be used as one of several decision criteria. A score that proves superior on a standalone basis may not be superior when incorporated into possible strategies.

In addition, do not evaluate score performance according to an isolated measure. A validation may present the effectiveness of a score based on one measure—K-S, for example—as if it is the only factor to consider. However, you should take into account how the score might perform when implemented in your environment.

Consider your operating threshold

When comparing two or more scores, validation results often indicate which score is superior across the entire distribution of accounts. However, you need to also evaluate the score in the operating areas of your organization, where the score will be used for decisioning. A score that performs well on your overall population may not necessarily be the one that is superior at your operating threshold, for example your score cutoff points.

Example

In this example, we examine the effectiveness of the FICO® Score using a bottom-line benefit analysis. The model is evaluated within the context of an originations strategy, including current acceptance criteria. We convert information from the validation along with the financial assumptions made for a lender's portfolio. When financial assumptions are not available specific to your portfolio, you can often leverage industry level figures as needed.

Currently, the company receives 20,000 applications a month, and accepts 70% of their applicants. The current strategy using their existing Score B to determine acceptable consumers, and achieves a bad rate of 8.5%. Using the FICO® Score, the lender can achieve a bad rate of 7.3% while maintaining the same 70% acceptance rate.

To determine the financial impact, this example further estimates the revenue and loss for each account (see Table 3). The company estimates that each bad account translates into a $2,500 loss. Given the FICO® Score increases the bad detection rate by 15%, the company is able to identify an incremental 183 bad accounts per month or 4,565 bad accounts per year; and thus, improving the decisioning process by not booking these accounts. The company further assumes each non-bad account translates to $300 lifetime revenue. However, given the improved rank ordering ability of the FICO® Score, the company is able to accept a similar number of non-bad accounts, thus maintaining similar revenue levels between the two strategies. Therefore, by using the FICO® Score in the company's originations strategy, the company would gain an incremental profit of $5.4M per year.

Metric	Current Strategy Using Score B	New Strategy Using FICO® Score
Acceptance Rate	70%	69%
Number of Applicants	20,000	20,000
Number of Accepts	14,000	13,800
Bad Rate	8.5%	7.3%
Number of Bad Accounts	1,190	1,007
Estimated Loss per Bad Account	$2,500	$2,500
Number of Non Bad Accounts	12,810	12,793
Average Revenue per Non Bad Account	$300	$300
Total Estimated Revenue	$3,843,000	$3,837,780
Total Estimated Losses	$2,975,000	$2,518,500
Total Estimated Profit	$868,000	$1,319,280
Profit Increase	**$451,280**	
Profit Increase (YEAR)	**$5,415,360**	

Table 3: Determining the financial impact of the FICO® Score

To cost-justify the FICO® Score, the company could then compare the cost of the model to the $5.4M per year incremental profit gained.

Other Considerations

There are several less quantitative, yet important considerations when validating and evaluating a score.

Performance over time

A validation should ideally consider multiple points in time to ensure that FICO® Score will continue to be robust and perform well over time. Therefore, if time and resources allow, consider conducting a retrospective analysis to validate and compare the FICO® Score performance at two or more points in time. In addition, once you have adopted the score, it is best practice to continue to track the FICO® Score's robustness regularly. The initial validation conducted on the FICO® Score can be established as a benchmark point for future comparisons.

Score interpretability

When comparing multiple scores, you also need to consider the interpretability of the scores. This is important for several reasons:

- **Legal compliance.** A key way to ensure legal compliance is to evaluate the ability of the score to accurately identify the reason for adverse action taken and quote logical reason codes.
- **Customer service.** In many applications, particularly in account originations, lenders are asked by consumers to explain their decisions. Reason codes that are easy to interpret can help you better explain to consumers why a certain action was taken.

Certain scores allow you to more easily understand the different characteristics within the score and the factors influencing the score. Reason codes and other interpretable features of the score allow for more targeted strategies and may help you respond to legal concerns, customer questions or internal inquiries.

For a full list of reason codes and how they are generated for the FICO® Score, please consult the FICO® Score User Guide. Additionally, www.scoreinfo.org explains how FICO® Scores are calculated and how consumers can manage their credit and their FICO® Scores over time.

Implementation costs

You should evaluate the cost of implementing a new score into your automated processing systems, as well as the cost of product maintenance. This may influence your decision on which score you select for your environment.

Implementation costs include your Information Systems group's time and resources to program the score or the interface to it. In addition, if score ranges are different from your previous score, standard monthly management reports will have to be reconfigured to accommodate the change. These issues can be examined with a cost-benefit analysis.

Appendix A: Glossary of terms

ACTIONABLE ACCOUNT/ACTIONABILITY:
An account where a lender, by taking action, can improve profitability. Within account management, for example, actionable bank accounts may be those who are not fully utilized and who are not yet charged-off or bankrupt; thus the issuer can improve profitability through line management, collections or marketing.

BAD(S):
Accounts with unsatisfactory performance during an observed performance period. For example, bads may be accounts that were 90+ days delinquent or worse during the performance period. The definition of “bad” is subjective and may vary from one organization or portfolio to another.

BAD RATE:
The fraction of the population that is considered bad.

CALIBRATE:
To benchmark the actual performance expected at each score range for a given portfolio. For example, a delinquency distribution report or retrospective validation may show that, of a group of accounts that scored 400–409, 1% went on to have a 30-day delinquency or worse in the subsequent year.

CUTOFF:
Any score value, designated by the user, above which and below which different decisions will be made on prospects, applicants or existing accounts.

DISTRIBUTION/SCORE DISTRIBUTION:
A table or plot of the percentage of a population (or population segment) by individual score or score range. Looking at a distribution of the total population could help determine volume impacts of different score cutoffs. Comparing the distributions of two separate performance groups (e.g. “goods” and “bads”) is also useful to evaluate the effectiveness of a score or model.

EXCLUSION CRITERIA:
Certain types of prospects, applicants or existing accounts that should not be included in a FICO® Score evaluation as their performance may be inaccurate or inappropriate (e.g. fraudulent account).

GOODS:
Accounts with satisfactory performance during an observed performance period. For example, goods might be accounts with no more than one incidence of 30-day delinquency during the performance period. The definition of “good” is subjective and may vary from one organization or portfolio to another.

INCIDENCE:
The occurrence of a particular account-related behavior of interest (e.g. bankruptcy, bad). An incidence-based model rank-orders consumers according to the likelihood of such an event occurring.

INDEPENDENT SAMPLE:
Accounts used to validate a model or score that were not used in the development of that model/score. An independent sample provides a more objective validation.

MODEL:
A statistical tool used for making decisions in credit risk management. A model is an algorithm that rank- orders individuals or accounts in a specific population according to a given outcome – e.g. likelihood of repayment.

OBSERVATION DATE:
In a validation, a past date or date window corresponding to the time at which a decision was made on a prospect, applicant or customer. Data would be taken from the observation date or a range of dates in a particular period of time (e.g. one quarter). This data would be compared with behavior in the performance period that followed.

ODDS:
Ratio of the number of accounts or individuals in one group to the number in an opposing group. For example, good:bad odds are the ratio of number of good accounts per every one bad account.

OPERATING THRESHOLD/REGION/RANGE:
The FICO® Score range in which major decisions are likely to be made – for example, the scores surrounding a FICO® Score cutoff.

PERFORMANCE DEFINITION:
The outcome of interest for a validation (e.g., “goods” vs. “bads”). The performance definition is the underlying logic for classifying accounts into these outcomes. To establish a performance definition, a pertinent question might include “How delinquent does an account need to be to be called a bad?”

PERFORMANCE WINDOW:
A set period of time following the scoring or observation date over which account performance is measured.

POPULATION:
The universe of prospects, applicants or existing accounts on which a credit grantor is making decisions. This may be a prospect base for prescreening, an applicant pool in account originations or a portfolio for account management decisions.

POPULATION SEGMENTATION:
The process of separating accounts into unique groups— known as subpopulations—that display common characteristics distinguishing them from other groups. In score validations, population segmentation is used to define groups of accounts on which to evaluate a score’s effectiveness. For example, a lender may want to test a score on accounts from a particular geographic region or on high revenue generators.

PREDICTIVE MEASURE:
A statistic (e.g., divergence, K-S, ROC) or plot (e.g., trade-off curve) that gauges the ability of a score/model to predict a future outcome.

PREDICTIVENESS:
The ability of a score/model to rank-order individuals according to a specific performance outcome. The more predictive a score/model is, the better it distinguishes between future “good” and “bad” performers.

RANK-ORDER:
The ability of a score or model to sort accounts according to their likelihood of exhibiting a specific outcome. For example, for the FICO® Score, strong rank-ordering is observed when accounts in each score range perform better than accounts scoring lower and worse than accounts scoring higher.

REASON CODE
Reasons returned along with a score that explain the most important factors influencing why the customer's score did not score higher (e.g., "Length of credit history is too short"). Reason codes are often used to provide "adverse action" reasons to consumers when they are denied credit, as required by both the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) in the U.S.

RETROSPECTIVE ANALYSIS:
The process of using archived data to test a model's predictive strength. After data for a specified performance period is collected, scores from a past point in time are appended to a group of accounts. The known observed performance on those accounts is then tabulated by score in order to evaluate how well the model predicted the actual outcome.

SCORE:
The numerical value generated by a model or group of models when applied to a given individual and/or account (e.g., 620).

SUBPOPULATION:
A distinct group of individuals within a given population that display common characteristics distinguishing them from other groups within the population.

SWAP SET:
The prospects, applicants or existing accounts for which a score will lead to a different decision than would otherwise be reached, such as by a lender's current decision process. For example, applicants currently declined by an existing score, but that the FICO® Score will show to be low-risk, or vice versa. As lenders evaluate scores to help them book and retain a more profitable set of accounts, swap sets may illustrate new opportunities for strategy refinement, such as rejecting more high-risk accounts and replacing them with lower-risk accounts.

THROUGH-THE-DOOR POPULATION:
The entire applicant population for a lender— both booked and non-booked applicants. Also referred to as the universe population in applicant screening.

TRUNCATED SAMPLE:
A sample of prospects, applicants or existing accounts that does not fully represent the population on which the score will be used and has an inherent "bias." For example, a sample consisting only of accounts that passed a lender's existing score cutoff— and not those that failed the cutoff—would be a truncated sample for a FICO® Score validation in the orginations decision context.

VALIDATION SAMPLE:
A random sample of prescreen prospects, applicants or existing accounts used to check a score's performance. Ideally this data was not used in the development of the model being evaluated.

Appendix B: Validation Checklist

The following questions provide an easy-to-use checklist that summarizes the various considerations you should think through when designing a FICO® Score evaluation and examining results. As you conduct the evaluation, you can reference the questions to ensure the results are effective and meaningful for your organization

Management Goals

- What are you trying to accomplish with the aid of the FICO® Score?
- Do any of these goals conflict?

Performance Definition

- The FICO® Score was developed to distinguish good vs. bad as measured by 90+ days past due or worse anytime over 24 month
- Your existing score was developed to distinguish:
- What definition(s) are of interest to you?
- Can you compute your performance definitions of interest on your database?
- Are all score comparisons based on identical parameters, as recommended?

Performance Window

- What are you trying to accomplish with the aid of the FICO® Score?
- For the population of interest, what is the optimal time duration over which the performance should be measured?
- Can you calculate the performance of concern strictly during the chosen performance window, as recommended (e.g., not including delinquency prior to the scoring date)?
- If relevant, how can you isolate “actionable accounts”?
- Are comparisons based on identical parameters, as recommended?

Validation Sample

- For all scores being compared, is the validation performed on the same sample, as recommended?
- Is the sample representative of your future population, as recommended?
- Was the existing score developed on the validation sample? If so, an independent sample should be chosen instead.
- Is there an adequate sample size, as recommended?

Population Segmentation

- What key segments are important to you?
- Can you identify these segments on your database?

- Do you have sufficient sample sizes available for each segment of interest (e.g., enough “bads”)?
- Are all comparisons based on identical parameters, as recommended?

Score Implementation Strategy

- What is the expected strategy that will be implemented with the FICO® Score?
- Based on this strategy, identify key operating threshold(s) (e.g., expressed in percentile of the total population)?
- Can you estimate:
  - Losses on a “bad” (account level or on average)?
  - Revenues on a “good” (account level or on average)?
  - Total costs (fixed and variable)?

Other Considerations

	Favors which Score?	
	FICO® Score	Existing Score
Expected performance over time		
Legal compliance		
Interpretable results		
Support and product maintenance		

Appendix C: Performance Data *Example* Layout

Field	Description/comments	Format	Length
Unique Identifier	Match to the output field with the same name in the CRA output file	Character	19
File Extract Date	This date corresponds to the date that the performance extract was created. Each record on a given file will have the same date. Must be in YYYYMMDD format	numeric	8
Performance Date	This date will either be account cycle date/due date or month end. Must be in YYYYMMDD format	numeric	8
Decision	Approve/Decline decision Indicator	Character	1
Override Reason	Reason account was booked/declined against recommendation of existing score/rules.	Character	2
Account Open Date	The date that the account opened. Must be in YYYYMMDD format	numeric	8
Account Close Date	The date that the account was closed. Must be in YYYYMMDD format	numeric	8
Account Closure Reason*	The reason that the account was closed (e.g. write-off, voluntarily closed, etc).	string	2
Maximum Delinquency**	The worst delinquency achieved on the account during the performance window	Character	1
Worst Account Status	The worst status achieved on the account during the performance window	Character	1
Ending Account Balance	The balance of the account at performance date or as of charge-off.	numeric	8
Ending Account Limit	The limit on the account at performance date or as of charge-off.	numeric	8
Portfolio	Indicates the portfolio of the account	Character	6

* The reason that the account was closed:

- I = Inactive (never active during the performance window)
- C = Closed (in good standing), usually by account holder request
- W = Write-off or other forced closure

** The performance variable that the lender will generate is equal to the Maximum Delinquency in the performance window, which ranges from the beginning of the performance window to the end of the performance window.

The performance variable can take on the following 1-digit alphanumeric values:

0 – The account has been current for the duration of the loan

1 – The account reached 1-cycle past due in the performance window

2 – The account reached 2-cycles past due in the performance window

3 – The account reached 3-cycles past due in the performance window

…

9 – The account reached 9+ cycles past due in the performance window

W – The account was written off in the performance window

<missing> – The account was never activated and therefore has no meaningful performance.

The following are examples of generating Maximum Delinquency:

Example 1: A/C No. - 123456, an account opened in August 2009

Month	Year	Current Status
Aug	2009	
Sep	2009	0
Oct	2009	0
Nov	2009	0
Dec	2009	0
Jan	2010	0
Feb	2010	0
Mar	2010	0
Apr	2010	0
May	2010	0
Jun	2010	0
Jul	2010	0
Aug	2010	0
Sep	2010	0
Oct	2010	0
Nov	2010	0
Dec	2010	0
Jan	2011	0
Feb	2011	0
Mar	2011	0

MaxDelq = 0

Example 2: A/C No. - 234567, an account opened in September 2009

Month	Year	Current Status
Sep	2009	
Oct	2009	0
Nov	2009	1
Dec	2009	2
Jan	2010	3
Feb	2010	4
Mar	2010	5
Apr	2010	5
May	2010	5
Jun	2010	W
Jul	2010	W
Aug	2010	W
Sep	2010	W
Oct	2010	W
Nov	2010	W
Dec	2010	W
Jan	2011	W
Feb	2011	W
Mar	2011	W

MaxDelq = W

Example 3: A/C No. - 345678, an account opened in April 2009

Month	Year	Current Status
Apr	2009	
May	2009	0
Jun	2009	0
Jul	2009	0
Aug	2009	0
Sep	2009	0
Oct	2009	0
Nov	2009	0
Dec	2009	0
Jan	2010	1
Feb	2010	2
Mar	2010	3
Apr	2010	0
May	2010	0
Jun	2010	0
Jul	2010	0
Aug	2010	0
Sep	2010	0
Oct	2010	0
Nov	2010	0
Dec	2010	1
Jan	2011	1
Feb	2011	2
Mar	2011	3

MaxDelq = 3