



Online Scoring and Reporting

# Uses for Validity Responses

---

Prepared by: Daisy Vickers  
V.P. Performance Scoring and Research

[www.mzdevinc.com](http://www.mzdevinc.com)  
[daisy.vickers@mzdevinc.com](mailto:daisy.vickers@mzdevinc.com)



---

# Uses for Validity Responses

Prepared by Daisy Vickers  
Vice President, Scoring & Research  
MZ Development

---

## Executive Summary

---

This paper discusses the use of validity responses to validate scoring accuracy as well as the uses for continued training, calibration, control of scorer drift and the use of OSCAR to make monitoring and intervention simple. Because the selection, administration, and scoring of validity responses is such a critical part of the scoring process, it is time consuming and expensive. The validation of the project and the fate of scorers can often hinge on the data. In order to be effective and efficient, it is important to get as much information as possible from the data and the responses.

## Validity Overview

---

Validity is the process of inserting into the pool of responses, pre-scored responses, not previously seen by scorers. The responses are distributed on a predetermined frequency throughout a project and are used to monitor accuracy and consistency in scoring.

True scores (defined as scores assigned by expert scorers) for these papers are loaded into the system and a report is run that shows what percentage of accuracy a scorer has achieved on validity papers in scoring against the true score. Ideally, these responses would be loaded at the beginning or before scoring begins to guard against the possibility of drift by the supervisor or expert scorer. Validity papers are used as a check to ensure that scorers, as well as scoring supervisors, are not drifting from the standards of the rubric, and are continuing to score accurately, or in a way that is valid based on the rubrics and training materials

While the validity of a scoring project is based on a number of statistics, the questions that often have to be answered at the end of a project are;

- Are the scores accurate?
- Did the scoring standards change during the project based on a scoring decision?
- Did the scorers apply the rubric consistently during the entire project?
- How can the scores be so different this year from past years?

While the Interrater Reliability Statistics (IRR) give feedback on how scorers are agreeing with one another, it is the validity statistics that give feedback on the accuracy of the scorers in applying the rubric based on the standards set by the client.

## OSCAR & Validity

---

OSCAR, a market leading distributed performance scoring platform, allows scoring directors the ability to assign responses for validity at any time before or during the project. Validity papers can be escalated from the current pool of responses or can be loaded from previous administrations. The process of selecting/loading responses for validity is simple.

OSCAR provides administrators on demand access to a series of validity reports. Validity reports can be viewed by item, by team, by individual scorer, by time period, or by score point. The reports provided are real-time and cumulative and can be seen in the form of statistics or by graphs. Frequency distribution from past years can be added to the reports so a quick glance can show drift across years or administrations. A snapshot of an early day's statistics can show drift within an administration. The reports can be generated by trait or holistically.

OSCAR gives control to scoring administrators to set a threshold of incorrect validity responses or a % of incorrect traits that a scorer can score inaccurately, and once the scorer reaches that

threshold, remediation (training or calibration) will need to occur before the scorer is allowed to score operational responses.

## Uses for Validity Responses

---

### Continued Training of Scorers

Because validity responses are scored by experts and the scores should not be questioned, some clients choose to use them to continue training of the scorers. The responses are usually solid score points, in other words, scorers should be able to score them correctly. When scorers score one incorrectly, it can be helpful for the scorer to be aware in order not to make the same mistake again. In projects with large scorer pools or distributed projects, it can be helpful to return the responses to the scorer with the correct score and an annotation for the score. This practice may increase the number of scorers a supervisor can monitor. Even scorers who scored the response correctly may benefit from the refresher. While this may be a waste of time for easy to score projects or for experienced, accurate scorers, it can certainly be a time saver for more difficult scoring projects.

OSCAR allows the administrator the ability to configure when validity feedback is presented to scorers. The following options are available:

- Feedback shown for only incorrect submissions
- Feedback shown for incorrect & correct submissions
- No feedback shown at all

### Identifying Scorer Drift

By analyzing the validity statistics by score point and pairing them with the frequency distribution report, the scoring supervisor should be able to identify scorer drift. If the frequency distribution is not in the expected range, the supervisor should look at the validity reports to see if the scorers are scoring high or low on each score point. This should be reviewed at the item level, team level, and individual scorer level.

OSCAR allows scoring supervisors the ability to address scorer drift by creating additional practice sets, delivering a calibration set, or removing the team/individual from operational scoring to complete additional training.

### Defining Calibration

If drift is detected in the analysis of the validity statistics, the supervisor may choose to use validity papers that posed a problem for scorers as calibration sets. If not, the responses and the information surrounding the scoring of them indicates what type of responses the scorers need to see for calibration. Because validity data at the score point level is readily available, moving the scorers from being consistently low or high to being accurate is possible. The calibration responses can be sent automatically at defined times or as needed.

OSCAR makes it simple for scoring supervisors to create calibration sets from a pool of responses that have been escalated from the operational pool or loaded from previous scoring administrations.

## Evaluating Scorers

Scorers are continuously evaluated using validity by a time range allows a scoring supervisor to monitor a scorer scoring today/this week and for an entire project. Sharing validity performance with a scorer is a helpful training tool. All of the statistics of a scorer are saved and can be printed out which makes conversations about performance concrete and actionable.

OSCAR gives administrators the ability to turn on validity reports for the scoring team. This allows individual scorers the ability to look at their validity performance by the same criteria as a scoring supervisor.

## OSCAR Customer Testimonial

---

*OSCAR is the first platform I have encountered in over 13 years in this business with the tools to take full advantage of the power of validity checks. Validity delivery and feedback can be completely customized and automated. And, since validity responses can also be selected using the normal scoring workflow, these features mean you can achieve any number of performance goals without stretching one of your most valuable resources, time.*

Tyson Watkins (Scoring Director) GCA

## For More Information

---

### About MZ Development

MZ Development is a dynamic and flexible organization focused on a single mission: to build a platform that delivers the most efficient and accurate solutions for performance assessment scoring and reporting. We enable our clients by providing cutting-edge technology to make performance scoring and reporting a simple experience, with rich results, in a more efficient manner than ever before.

MZ Dev developed OSCAR, a state-of-the-art, next-generation platform for online scoring and reporting that boasts a set of features to accommodate a variety of performance scoring requirements, while still achieving the primary goal of providing an accurate, simple, affordable solution. ([www.mzdevinc.com](http://www.mzdevinc.com))

For more information, please contact [daisy.vickers@mzdevinc.com](mailto:daisy.vickers@mzdevinc.com)

© 2018 MZ Development • All Rights Reserved