# Article information:

Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability - Sara Cushing Weigle, 2010
<https://journals.sagepub.com/doi/10.1177/0265532210364406>

# Article summary:

1. The study compares human and automated scores on responses to TOEFL iBT Independent writing tasks with non-test indicators of writing ability.

2. Automated scores were produced using e-rater, developed by Educational Testing Service (ETS).

3. Correlations between both human and e-rater scores and non-test indicators were moderate but consistent, providing criterion-related validity evidence for the use of e-rater along with human scores.

# Article rating:

May be slightly imbalanced: The article presents the information in a generally reliable way, but there are minor points of consideration that could be explored further or claims that are not fully backed by appropriate evidence. Some perspectives may also be omitted, and you are encouraged to use the research topics section to explore the topic further.

# Article analysis:

The article "Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability" by Sara Cushing Weigle presents a study that compares human and automated scores on responses to TOEFL® iBT Independent writing tasks with several non-test indicators of writing ability. The study aims to provide criterion-related validity evidence for the use of e-rater along with human scores.

The article is well-structured, with a clear introduction, methodology, results, and discussion sections. The author provides a comprehensive review of the literature on automated scoring and its potential benefits and limitations. However, there are some potential biases in the article that need to be addressed.

Firstly, the author only uses non-test indicators of writing ability as criteria for validation. While these indicators are useful, they may not be sufficient to validate the use of automated scoring in high-stakes testing situations. For example, the author does not consider how well automated scoring performs in predicting academic success or job performance.

Secondly, the author only uses one type of automated scoring system (e-rater) developed by Educational Testing Service (ETS). While e-rater is widely used in educational settings, it is not the only system available. Therefore, the results may not be generalizable to other systems.

Thirdly, while the author acknowledges some limitations of automated scoring systems such as their inability to detect sarcasm or irony, she does not explore other potential risks associated with their use. For example, there is a risk that students may learn how to game the system by using certain keywords or phrases that are known to increase their scores.

Finally, while the author discusses some implications of her findings for the validity of automated scores, she does not explore counterarguments or alternative interpretations. For example, it is possible that correlations between human and e-rater scores were moderate because both were influenced by similar factors such as grammar and syntax rather than because e-rater is a valid measure of writing ability.

In conclusion, while the article provides valuable insights into the validity of automated scoring systems, it has some potential biases and limitations that need to be addressed. Future research should consider using multiple types of automated scoring systems and criteria for validation to provide a more comprehensive understanding of their potential benefits and limitations.

# Topics for further research:

* Validity of automated scoring systems in predicting academic success
* Job performance and automated scoring systems
* Comparison of different types of automated scoring systems
* Risks associated with the use of automated scoring systems
* Gaming the system in automated scoring
* Alternative interpretations of the correlation between human and automated scores

# Report location:

<https://www.fullpicture.app/item/7a7f00d023f1192db1f9beeff461e01d>