



A Comparison of Anchor Lengths and Item Selection Methods in Small-Sample Equating

Kylie N. Gorney

University of Wisconsin-Madison

National Council on Measurement in Education (NCME)
Annual Meeting · 2021



Background

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- Several equating methods use a set of common, anchor items
- Historically, it has been recommended that anchor items be representative of the total test in terms of content and statistical characteristics (Kolen & Brennan, 2014)
- Could be challenging in practice, relaxing certain requirements may be beneficial (Sinharay & Holland, 2006)



Purpose

Anchor Lengths and Item Selection Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

The purpose of this study was to compare the effects of using different anchor lengths and item selection methods in small-sample equating.

- Fully crossed factors
- Treated the difference in test form difficulties as a random effect
- Treated the difference in group abilities as a random effect



Data

Anchor Lengths and Item Selection Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- Simulated a 100-item test
- Anchor lengths: $V = 20, 25, 30, 35, 40$
- Sample sizes: $N = 10, 25, 50, 100$
- 1,000 replications



Anchor Item Selection Methods

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

Requirements

- The anchor item p -values could not be 0 or 1
- The mean of the anchor test p -values had to be within 0.01 of the mean of the old form p -values
- The standard deviation of the anchor test p -values had to be within 0.01 of a certain percentage of the standard deviation of the old form p -values

Anchor item selection methods

- ① Minitest (100%)
- ② Semi-miditest (90%)
- ③ Miditest (75%)



Equating Design and Methods

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- Common-item nonequivalent groups design
- Equating methods
 - ① Identity (Id)
 - ② Synthetic (Syn)
 - ③ Nominal weights mean (NM)
 - ④ Circle-arc (Cir)



Bias

Anchor Lengths and Item Selection Methods

Gorney (2021)

Background

Purpose

Method

Results & Discussion

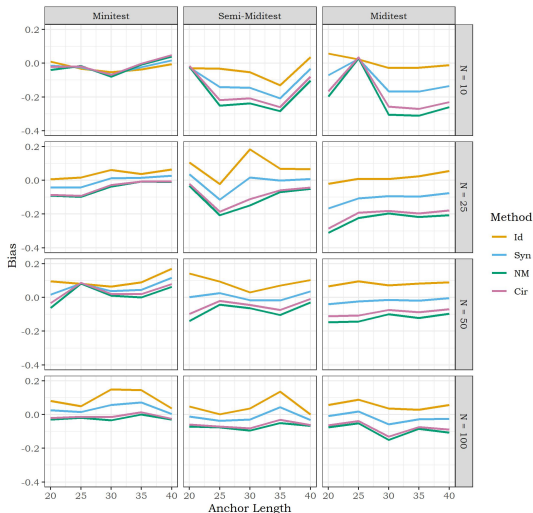
Bias

Root Mean Squared Error (RMSE)

Conclusion

References

- The minitest tended to produce the least biased results, while the miditest tended to produce the most biased results
- Anchor length had little to no impact on the average bias, though longer anchors tended to produce more stable results





Bias

Anchor Lengths and Item Selection Methods

Gorney (2021)

Background

Purpose

Method

Results & Discussion

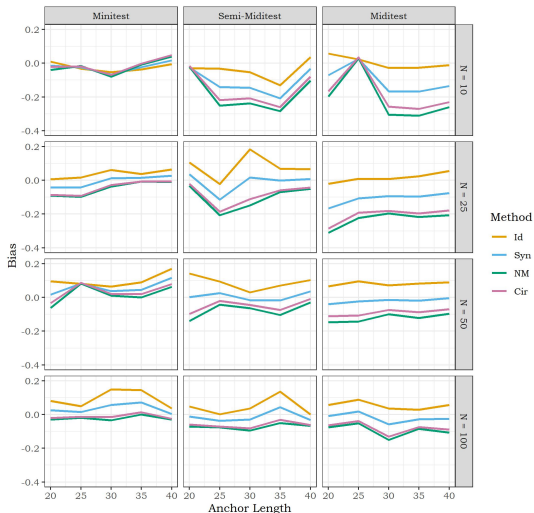
Bias

Root Mean Squared Error (RMSE)

Conclusion

References

- Identity equating yielded the most inconsistent amounts of bias
- Nominal weights mean equating and circle-arc equating produced most biased results, on average





Root Mean Squared Error (RMSE)

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

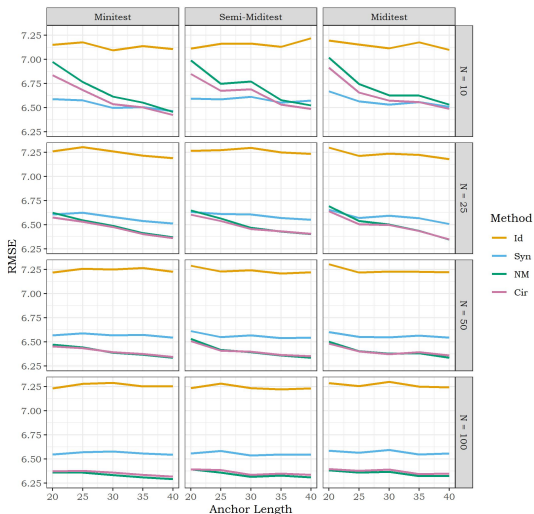
Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- The minitest produced slightly smaller RMSEs, on average, when the sample size was 10
- All item selection methods performed similarly for larger sample sizes
- Longer anchors were associated with smaller and more stable RMSEs





Root Mean Squared Error (RMSE)

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

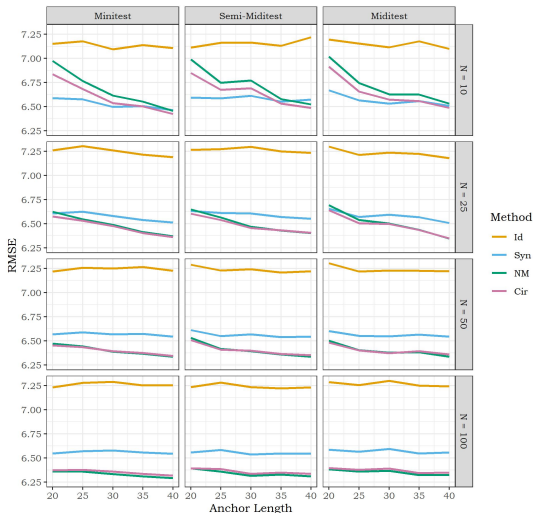
Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- Identity equating produced the largest and most inconsistent RMSEs
- Synthetic equating performed well when the sample size was 10
- Nominal weights mean equating and circle-arc equating preferred for larger sample sizes





Conclusion

Anchor Lengths and Item Selection Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- The minitest tended to produce the most accurate results, followed by the semi-miditest, and then the miditest
- Differences between anchor item selection methods decreased as sample size increased
- The use of longer anchors tended to decrease the variability of the results
- Nominal weights mean equating and circle-arc equating performed well across most conditions
- Future research
 - Simulation studies that use different anchor item selection methods and equating methods
 - Real data examples



Thank you!

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

For more information, please contact kyliengorney@gmail.com.



References

Anchor Lengths
and Item
Selection
Methods

Gorney (2021)

Background

Purpose

Method

Results &
Discussion

Bias

Root Mean Squared
Error (RMSE)

Conclusion

References

- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer.
<https://doi.org/10.1007/978-1-4939-0317-7>
- Sinharay, S., & Holland, P. (2006). *The correlation between the scores of a test and an anchor test* (RR-06-04). Educational Testing Service.