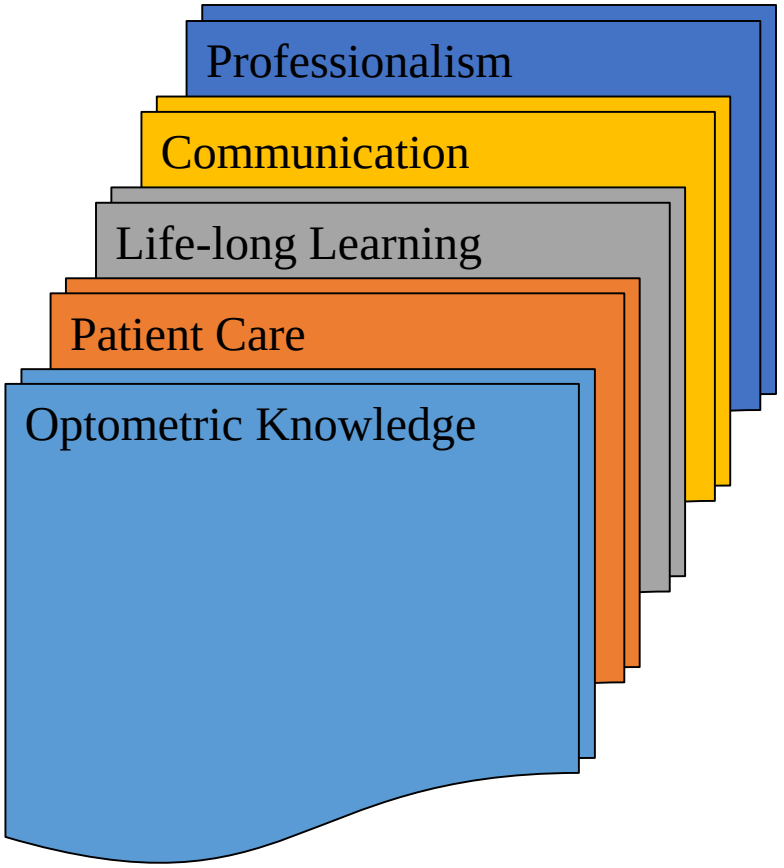
The accuracy of standard setting using the borderline regression method for varying cohort sizes

Dr Michael Pianta
Department of Optometry & Vision Sciences
MSHS, FMDHS
The University of Melbourne

The context

- Portfolio assessment
  - Used in the final year of the Doctor of Optometry
  - Students collate evidence that they have met leaning outcomes in five key areas
  - Each area is marked independently
    - Checklist score
    - Global rating
  - Students must pass all five areas to pass the year



The importance of standard setting

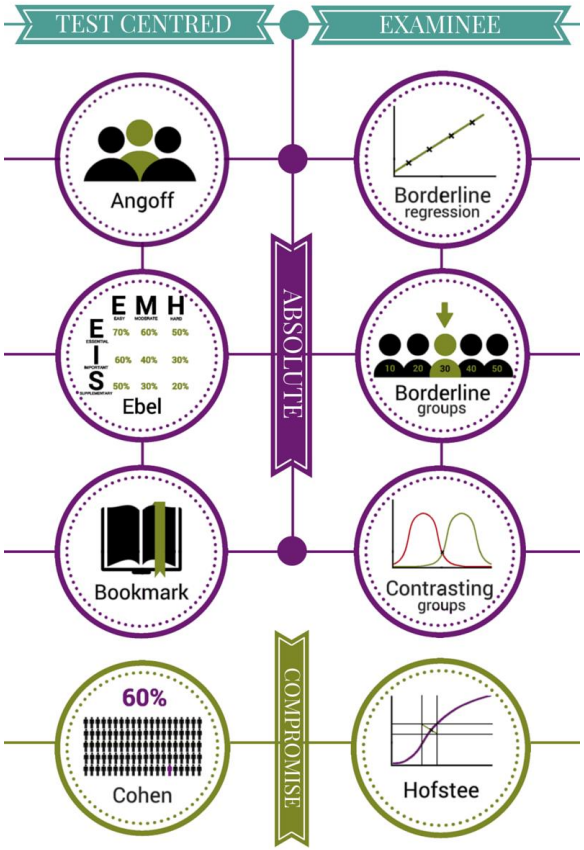
- Used to distinguish between a competent and incompetent student
  - Cut score
- Particularly important for high-stakes assessment
- Must be credible
  (Norcini & Shea, 1997; Norcini, 2003)
  - Defensible
  - Supported by evidence
  - Feasible
  - Acceptable to all stakeholders

http://opentuition.com/articles/acca-articles/how_to_pass_acca_exams/

Borderline regression

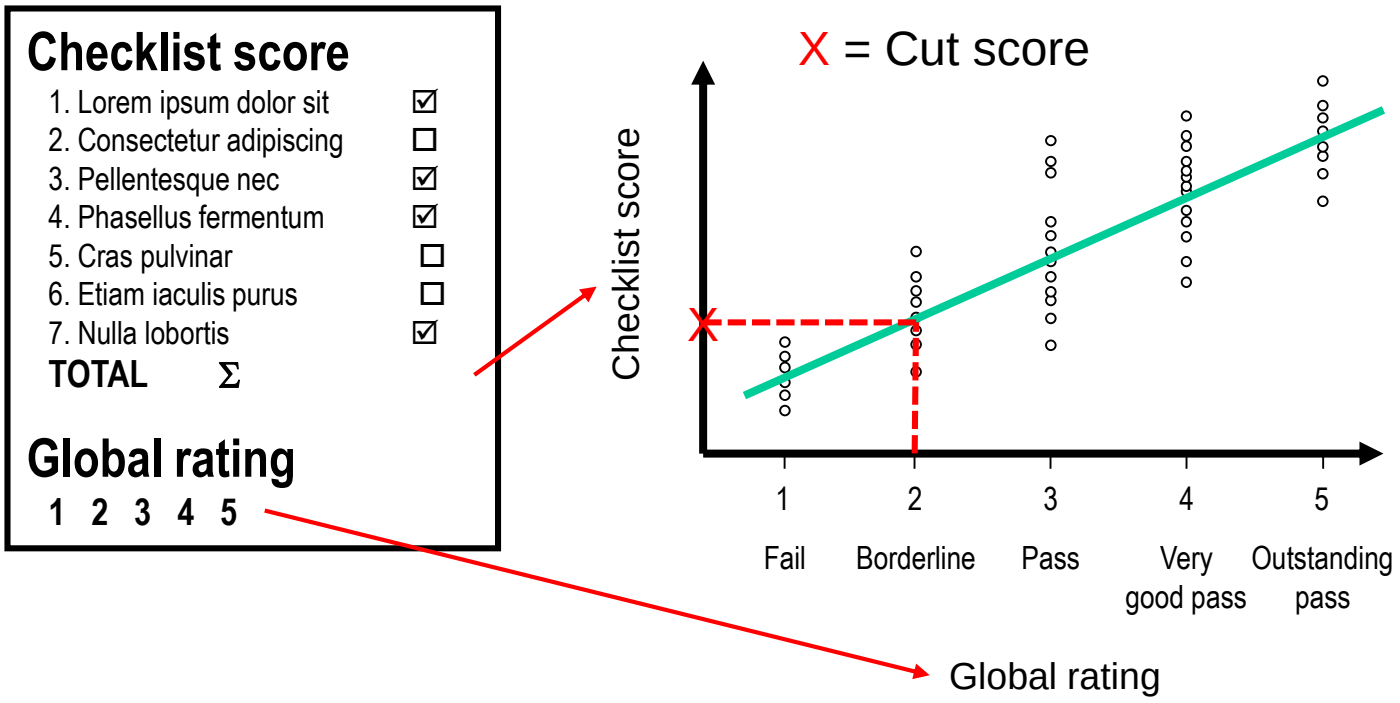
- Criterion-referenced (absolute)
- Examinee-centered
- Method
  - Checklist scores regressed on global ratings
  - Linear equation used to calculate the checklist score that corresponds to a global rating of "borderline" (the cut score)

(Kramer et al, 2003; Woehr & Fehrmann, 1991)



https://www.maxinity.co.uk/blog/2016/6/2/standard-setting-simplified-assessments

Borderline regression

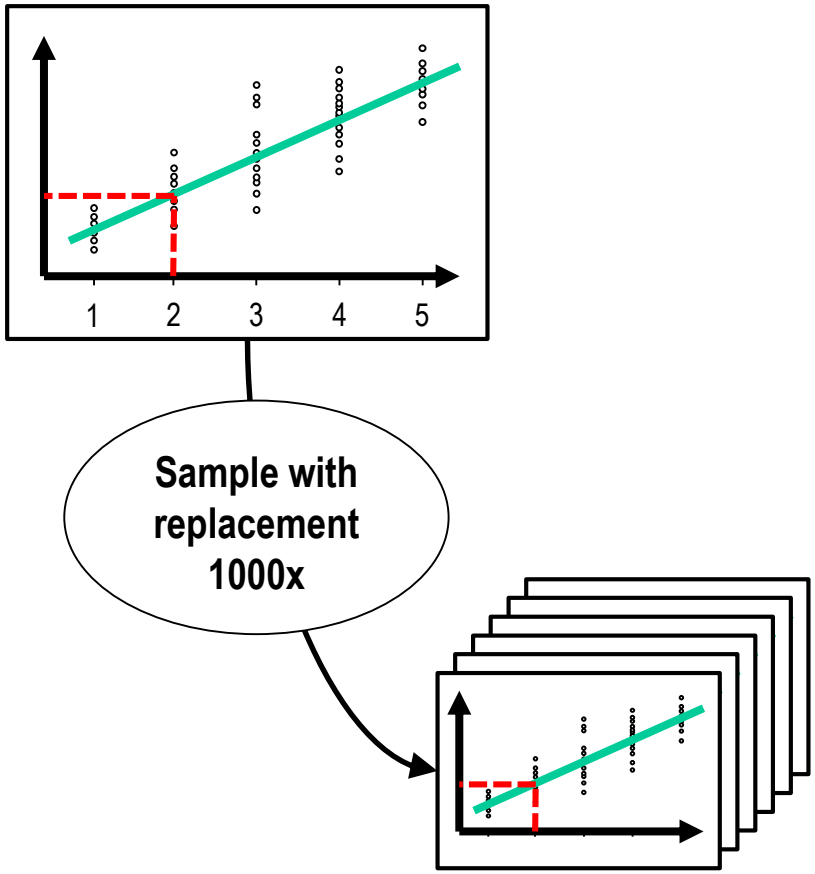


Aim

- To investigate how the accuracy of cut score estimates obtained using the borderline regression method varies with cohort size

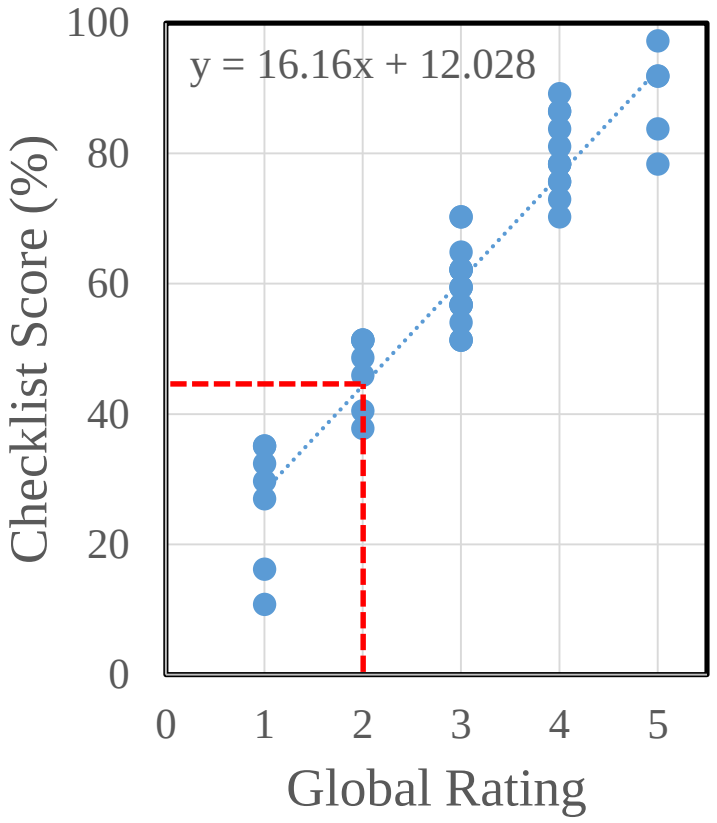
Methods

- Data
  - Data from final year portfolio assessment in the Doctor of Optometry (n=49)
  - Five learning outcome areas
- Bootstrapping (resampling) used to estimate standard error in the
  - Cut score
  - Coefficient of determination (R^2)
  - Cronbach's alpha (not shown)
- Simulated cohort sizes from 15 to 480 candidates

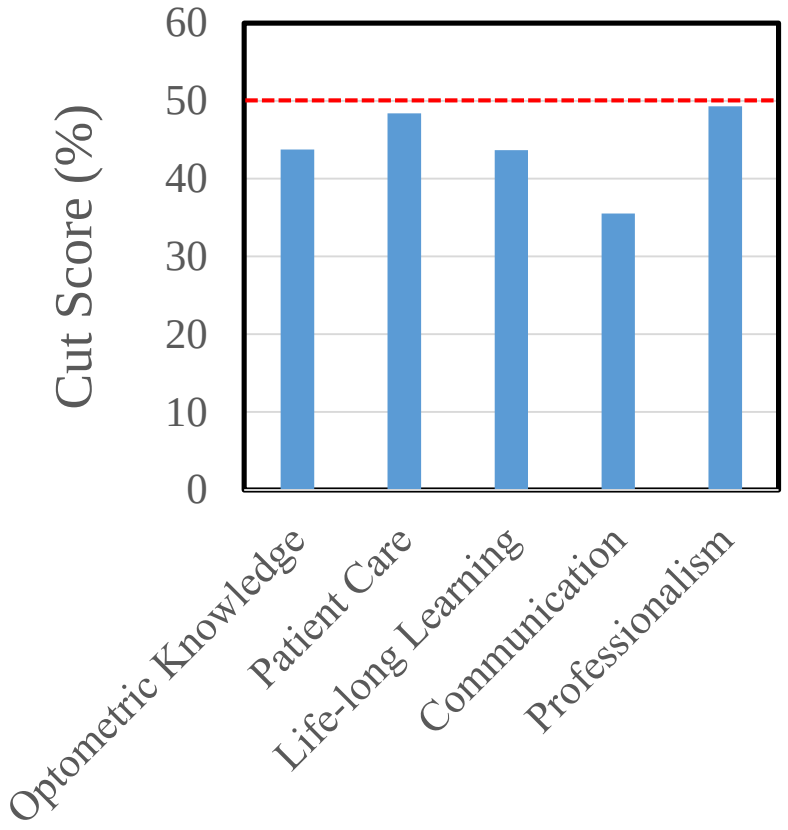


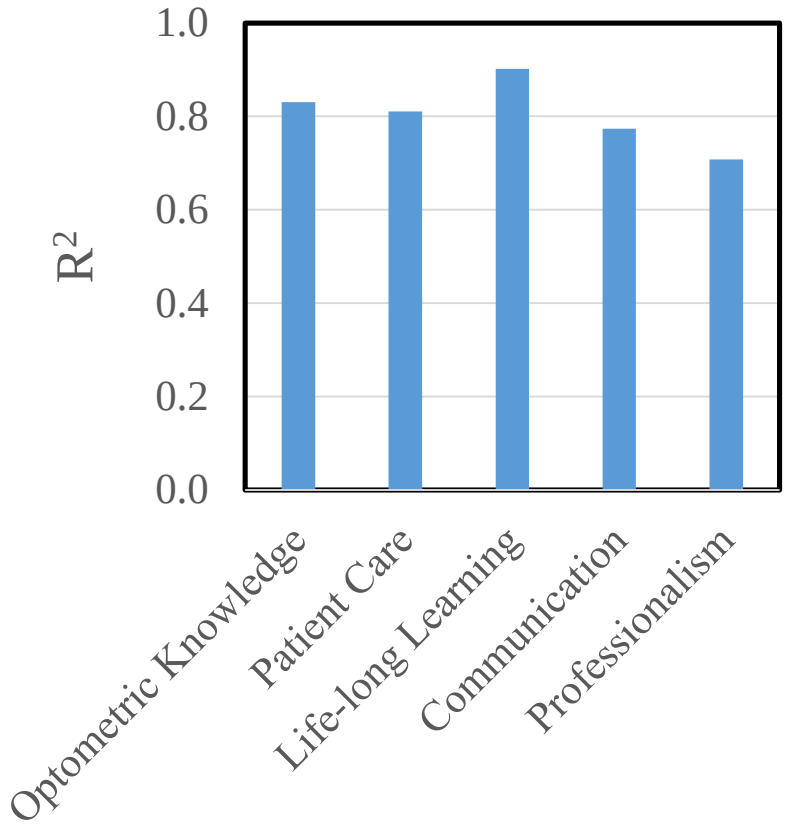
Sample borderline regression data

- Life-long learning area
  - N = 49
  - Cut score = 44.3%
  - R^2 = 0.89

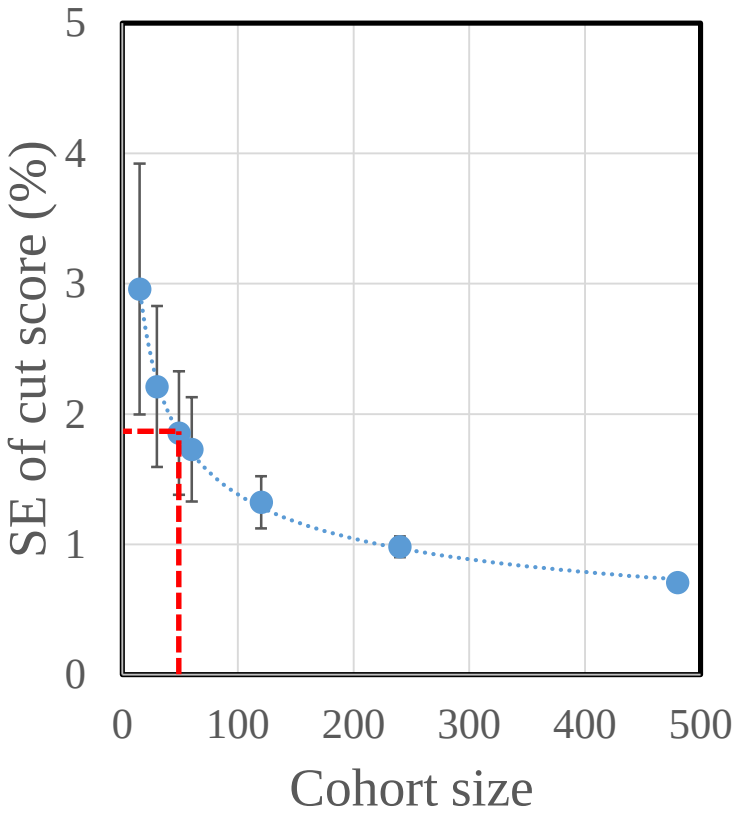


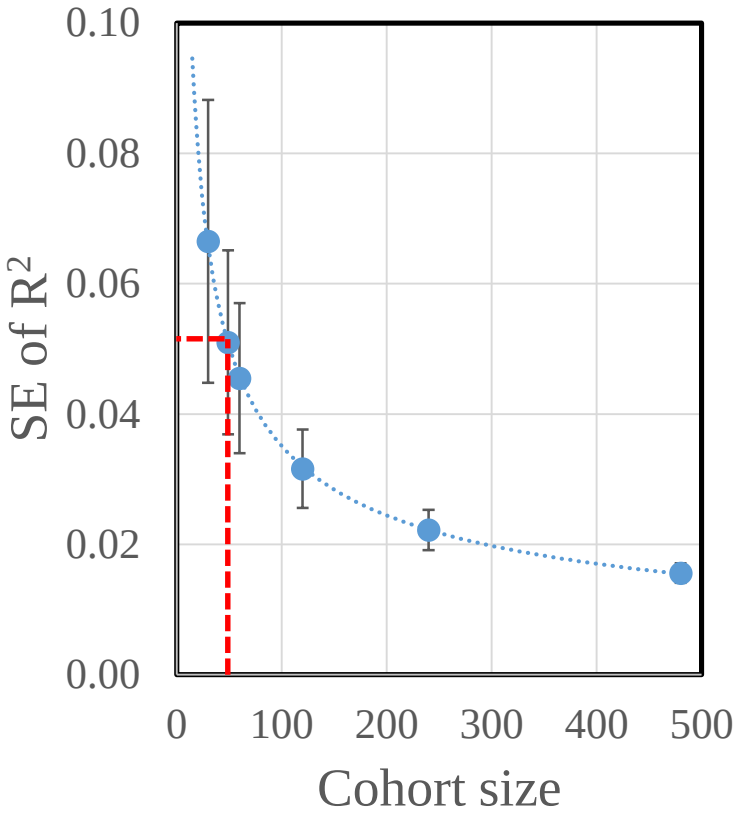
Cut scores and R^2 for real cohort





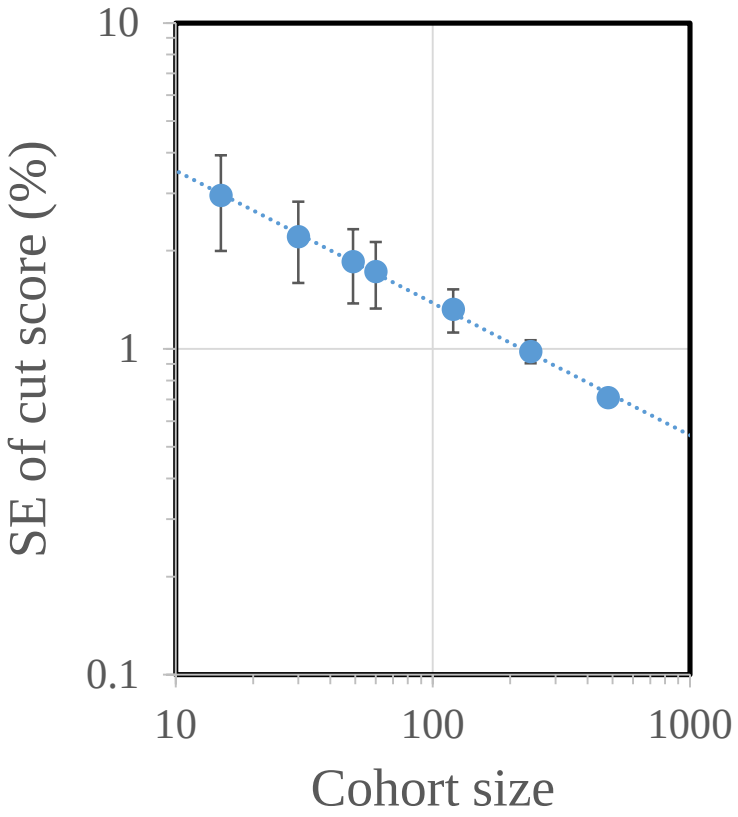
Standard errors for cut score and R^2

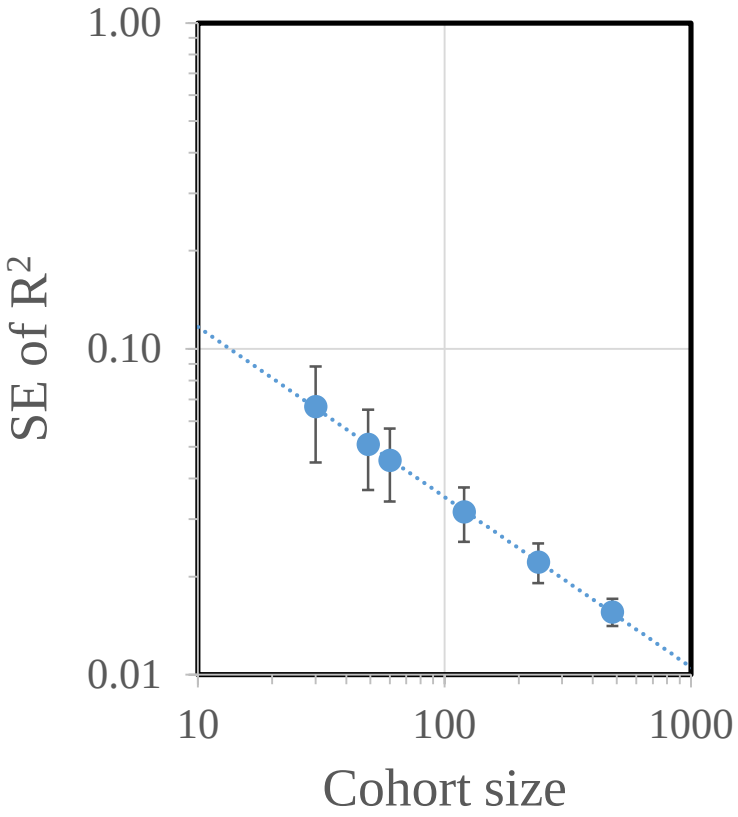




$y = ax^{-b}$

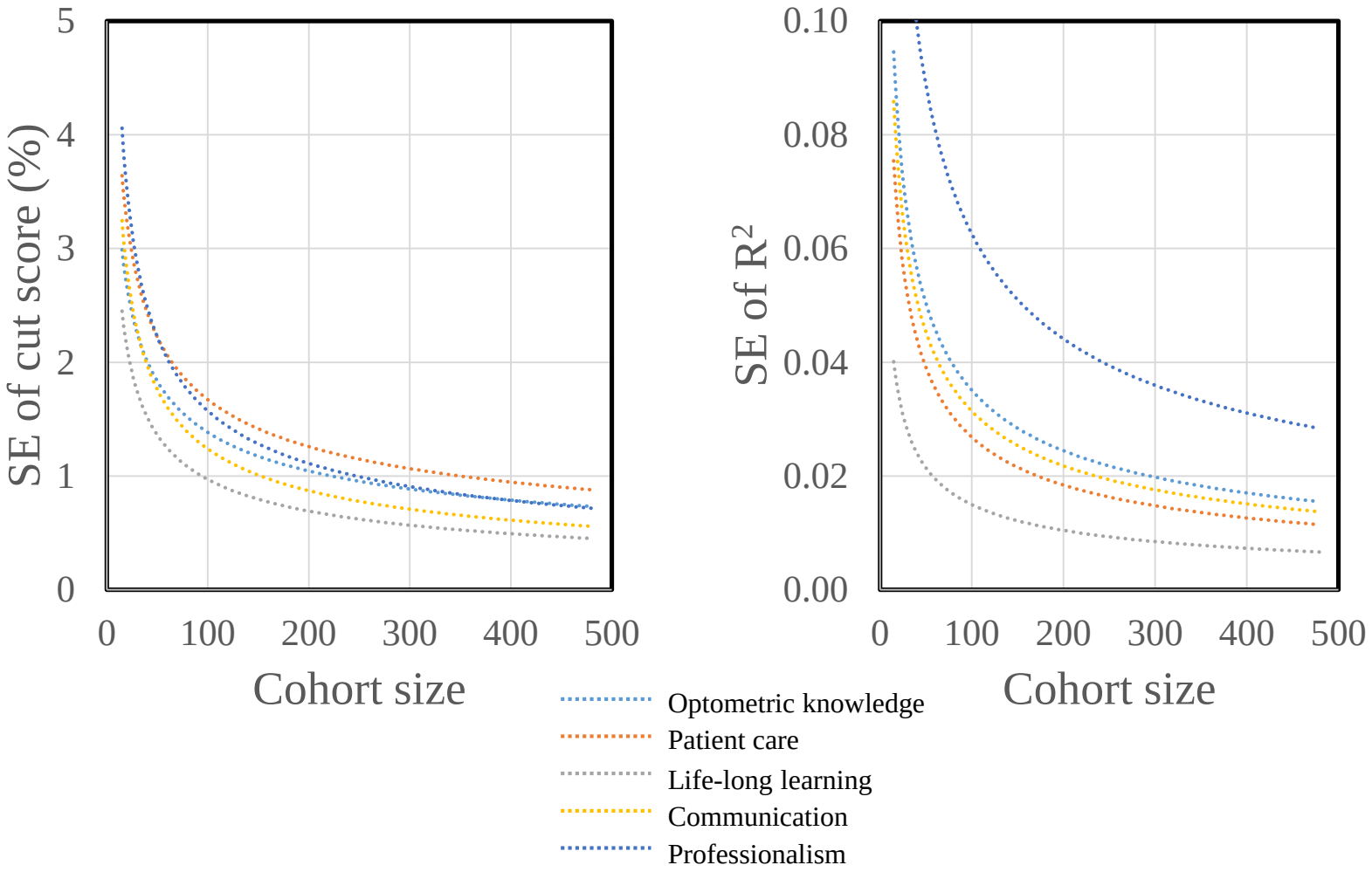
Standard errors for cut score and R^2



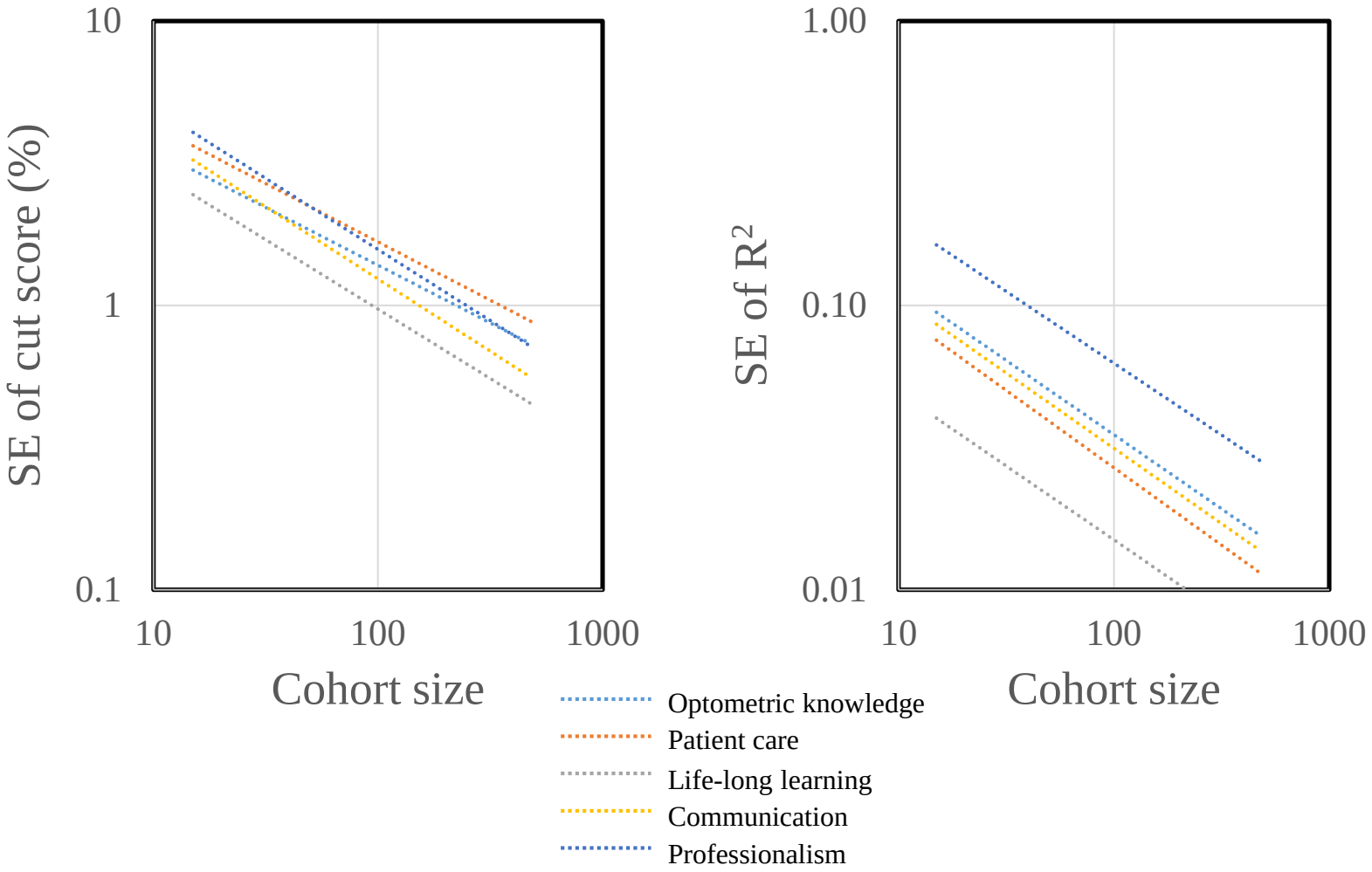


$$\log(y) = -b \cdot \log(x) + \log(a)$$

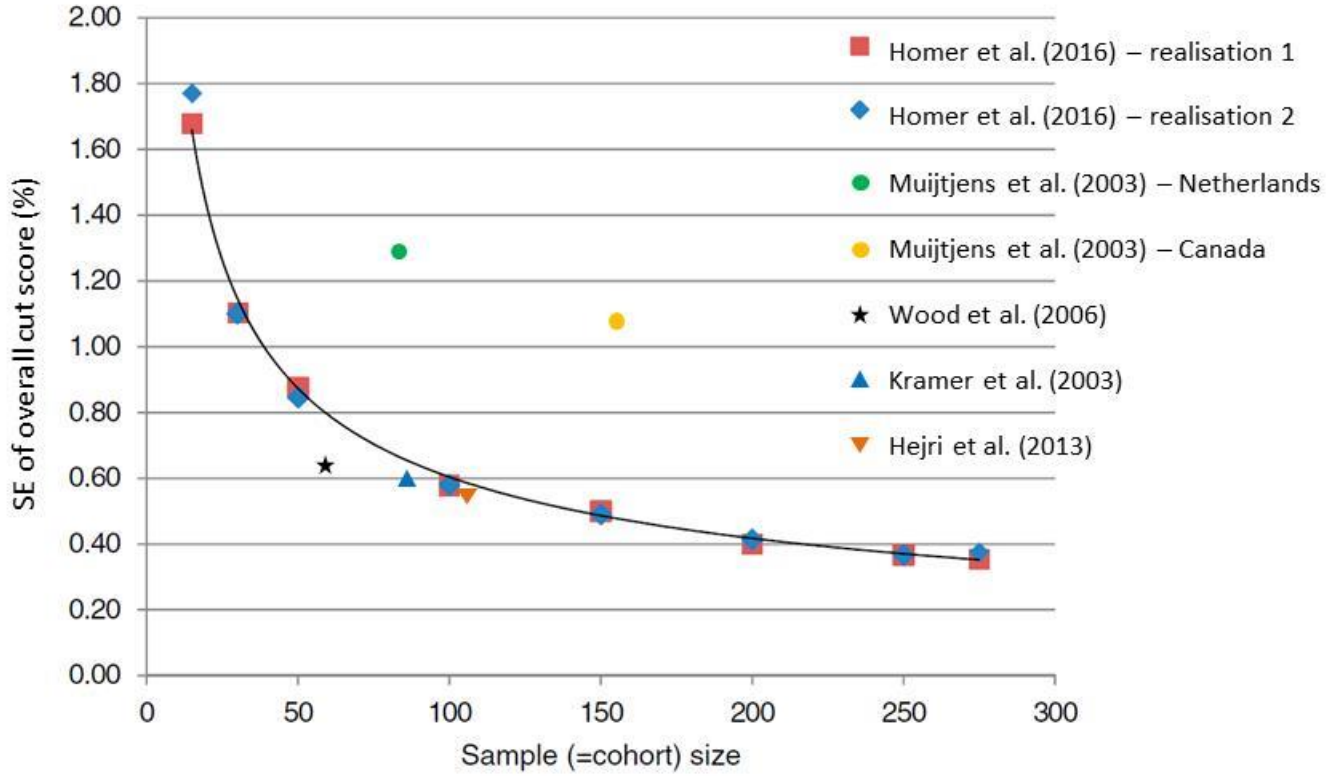
Standard errors for cut score and R^2



Standard errors for cut score and R^2

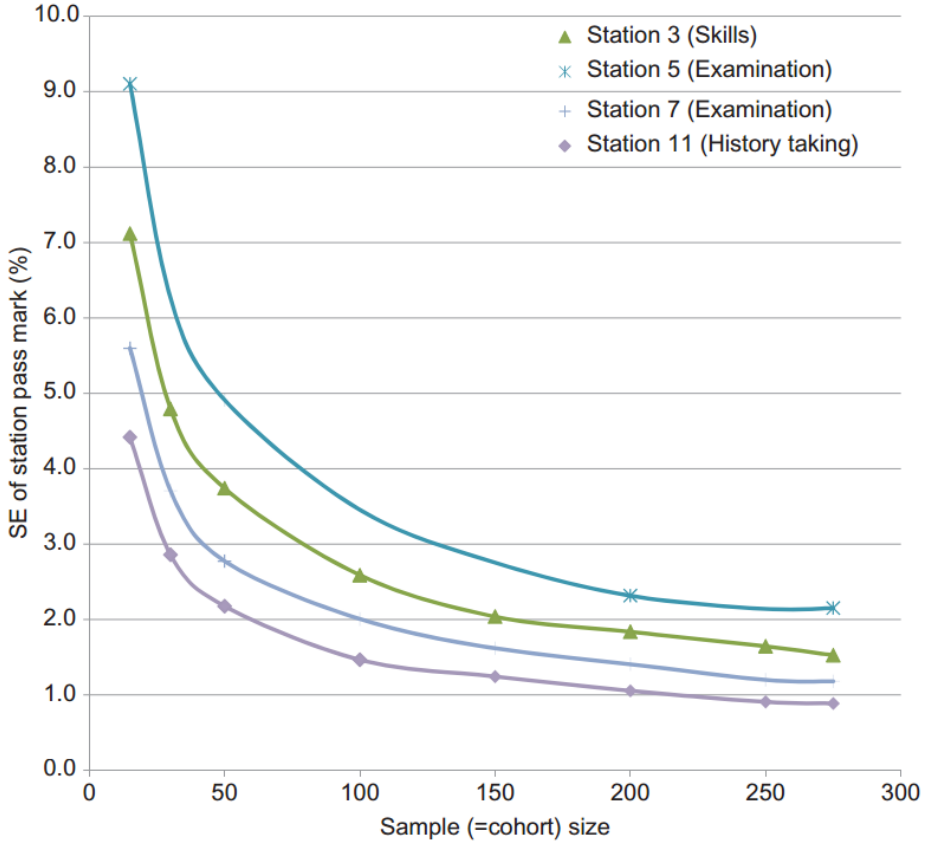


Comparison to previous work



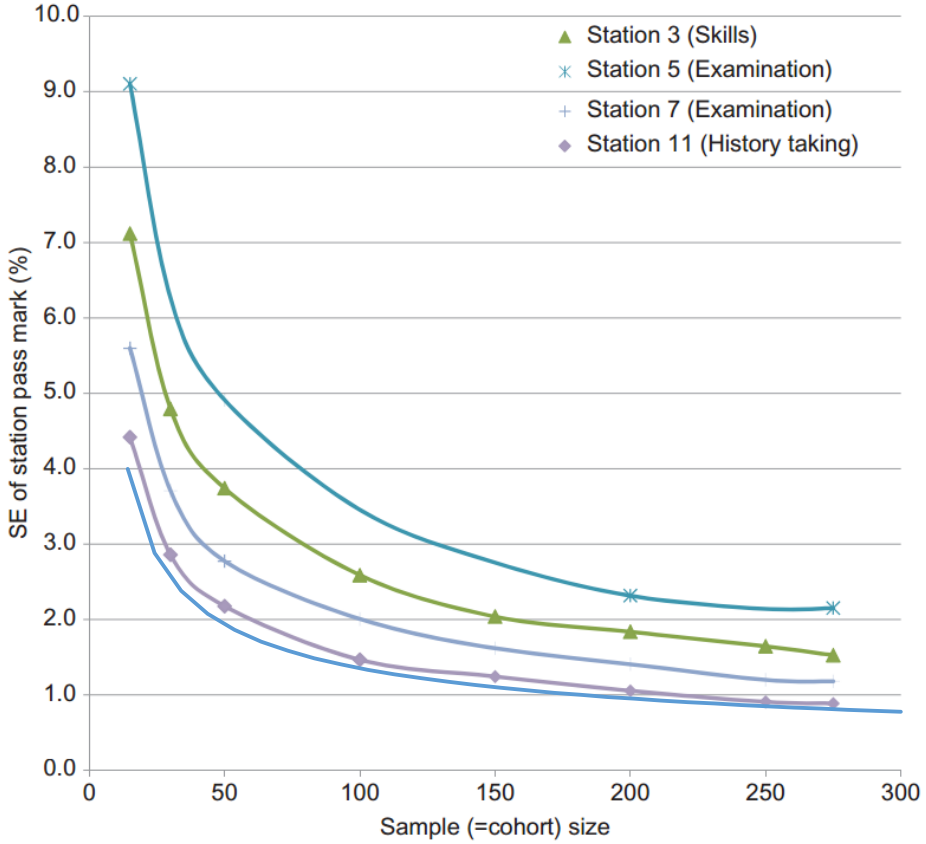
Homer et al (2016). *Med Teach*, 38: 181-188

Comparison to previous work



Homer et al (2016). *Med Teach*, 38: 181-188

Comparison to previous work



Homer et al (2016). *Med Teach*, 38: 181-188

Conclusions

- Bootstrapping can be used to calculate the SE in the cut score calculation for examinee-centered methods of standard setting (e.g. borderline regression)
- SE decreases as cohort size increases
  - Inverse square-root power law
- SE for OD cohorts ($n \approx 60$) is less than ≈2% for all portfolio learning outcome areas
- Bootstrapping is a simple and robust method for understanding the accuracy of standard setting, which can inform quality assurance for assessment