

Psychometrika at 90: Past, Present, and Future

Presidential Panel Discussion
IMPS 2026
Yonsei University
July 23, 2026

March, 1936

First Issue of Psychometrika

Psychometrika

A JOURNAL DEVOTED TO THE DEVELOPMENT OF PSYCHOLOGY AS A QUANTITATIVE RATIONAL SCIENCE



“Paul Horst offered to underwrite the losses of the journal for the first year for up to one-fourth its cost.”

ANNOUNCEMENT

STUDENT MEMBERSHIP AND SUBSCRIPTION

It was voted at the September, 1937 meeting of the Psychometric Society to provide for student memberships, including a subscription to Psychometrika, at the rate of \$7.00 per year. The regular rate for full membership is \$3.00 per year. The student member has all the privileges of a full member of the Society, except the right to vote. A person may hold a

Organized by

Sandip Sinharay, ETS (Executive Editor)

Ji Seung Yang, University of Maryland (Review Section Editor)

Participants (Speakers and Panel Discussants)

Dan Bolt, University of Wisconsin–Madison

Yunxiao Chen, London School of Economics and Political Science

Alison Cheng, University of Notre Dame

Matthew Johnson, ETS

Inhan Kang, Yonsei University

Stella Kim, University of North Carolina at Charlotte

Klaas Sijtsma, Tilburg University

David Thissen, University of North Carolina at Chapel Hill

A Question for All

How will psychometric research be disseminated in 20 years?

- A. Mostly through traditional journals
- B. Mostly through open repositories with post-publication review
- C. A hybrid of journals and repositories
- D. AI-curated living documents
- E. Something completely different



Three Perspectives

A brief “exclusive” history of Psychometrika - **David Thissen**

The 5 most seminal Psychometrika papers - **Klaas Sijtsma**

Where we’re now and where we’re going - **Yunxiao Chen**

“Provocations”

1. How should Psychometrika adapt to AI? (Alison Cheng)
2. What would an ideal peer-review system look like? (Dan Bolt)
3. Should the name of “Psychometrika” be replaced by something else that more appropriately reflects the current focus of the journal? (Inhan Kang)
4. Should we be reporting measurement results on a latent scale? (Matthew Johnson)
5. Should Psychometrika focus less on methodology and more on applications? (Stella Kim)

Panel Discussion (Round I&II)

- Should journals exist at all, or should people just upload their manuscripts on open-source platforms? (Klaas Sijtsma)
- “What is one idea from the past 90 years that won’t survive the next 20?” (David Thissen)
- “What should Psychometrika publish more of?” (Yunxiao Chen)
- “Should the 3PL model be laid to rest?” (Matthew Johnson)
- “How can Psychometrika attract more readers and authors from outside Europe and North America?” (Inhan Kang)
- In 20 years, what will the journal article look like? Will it still be a PDF? (Stella Kim)
- How can Psychometrika remain the flagship journal of psychometrics over the next 20 years? (Alison Cheng)
- What is the most important unsolved problem in psychometrics? (Dan Bolt)

Panel Discussion (Q&A)

“One prediction for 2036 Psychometrika”

Survey Result

A Brief History of
Psychometrika

David Thissen

with the (unwitting) assistance of

Willem Heiser, Paul Horst, Larry Hubert, Lyle Jones,
Sophia Rabe-Hesketh, & John Stalnaker

“The Founders”



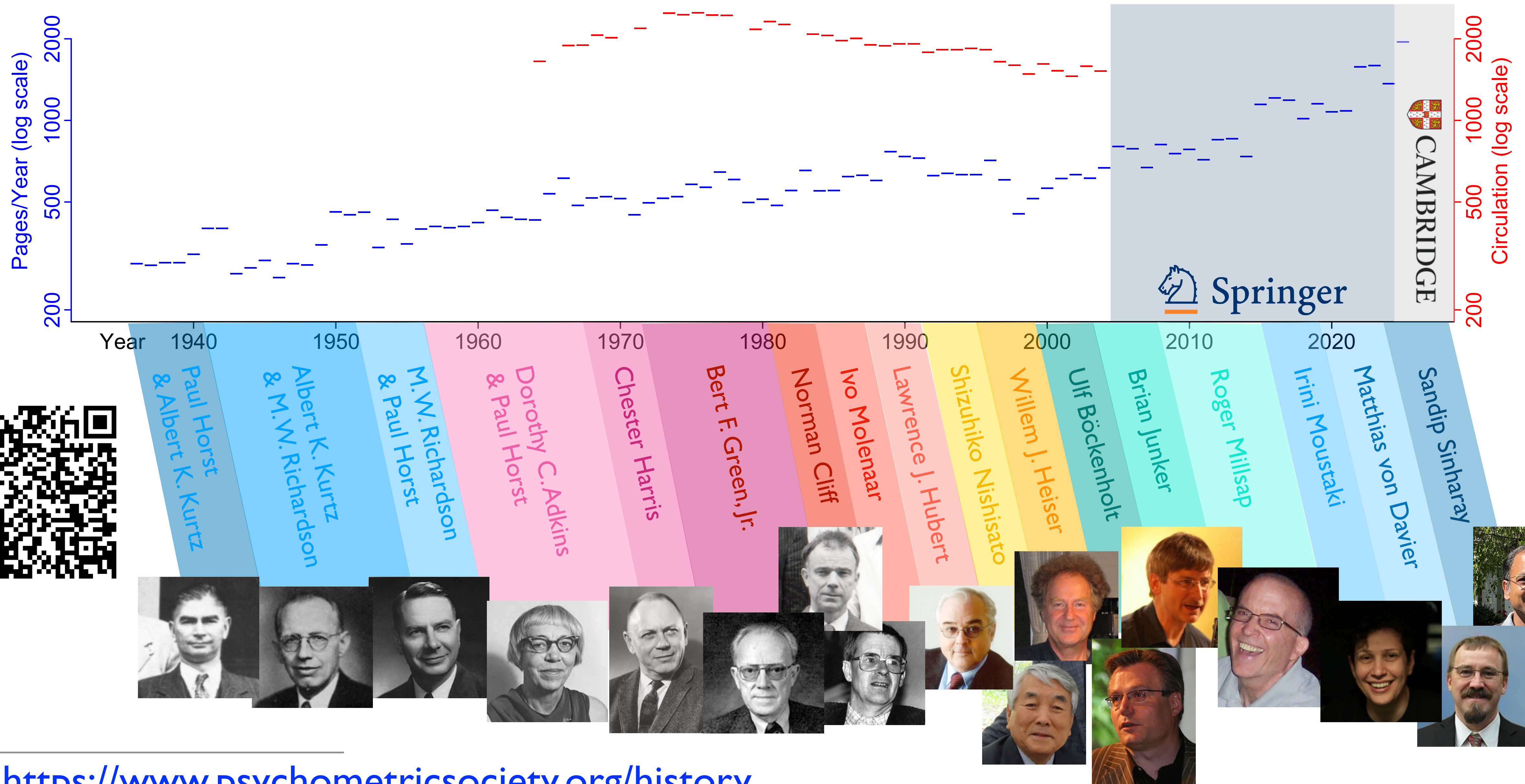
(from left to right) Albert K. Kurtz, Paul Horst, L. L. Thurstone, John Stalnaker, Marion W. Richardson and Jack W. Dunlap

“A **JOURNAL** DEVOTED TO THE
DEVELOPMENT OF PSYCHOLOGY AS
A QUANTITATIVE RATIONAL SCIENCE”

<https://www.psychometricsociety.org/history>

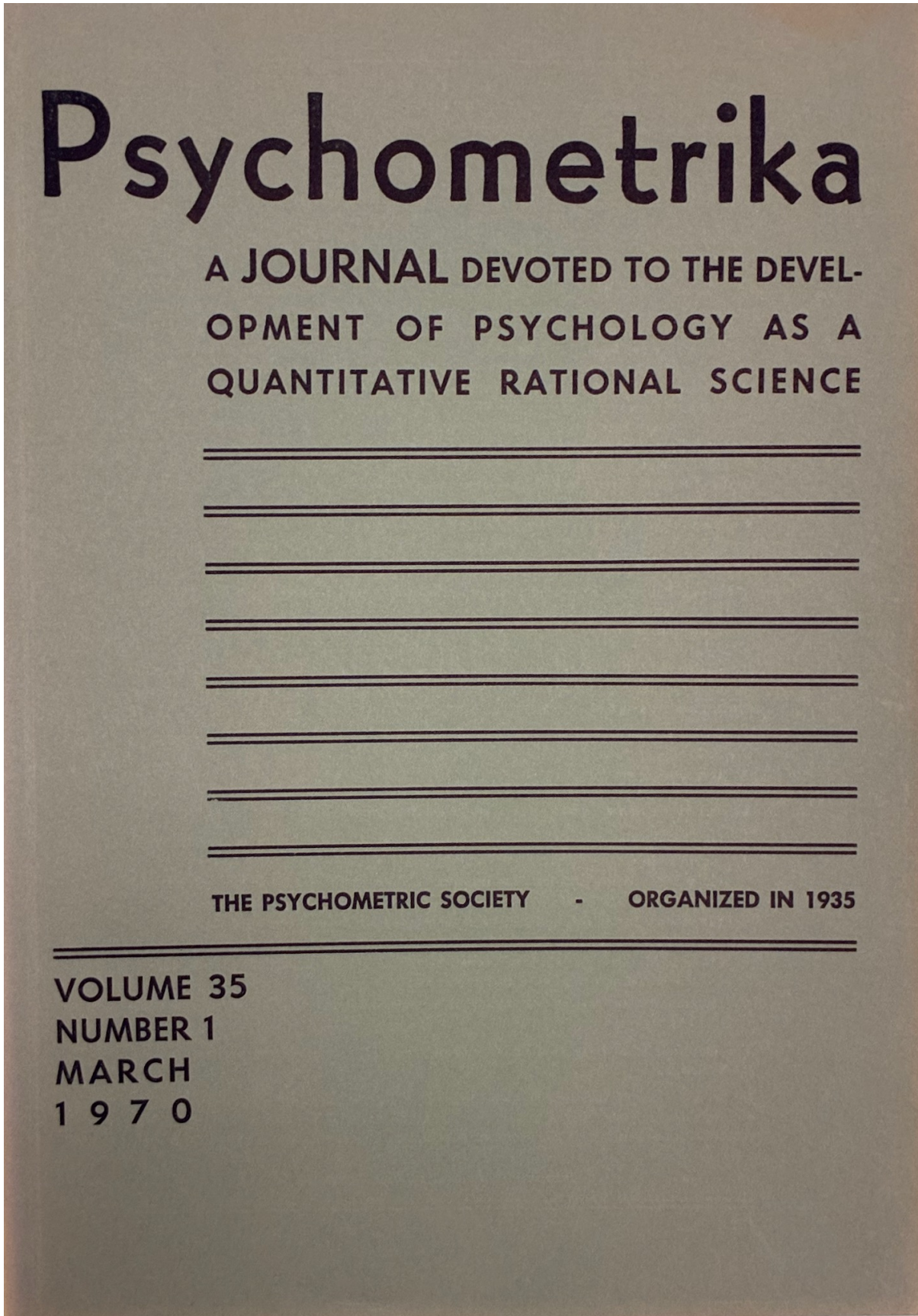


Psychometrika over the Years

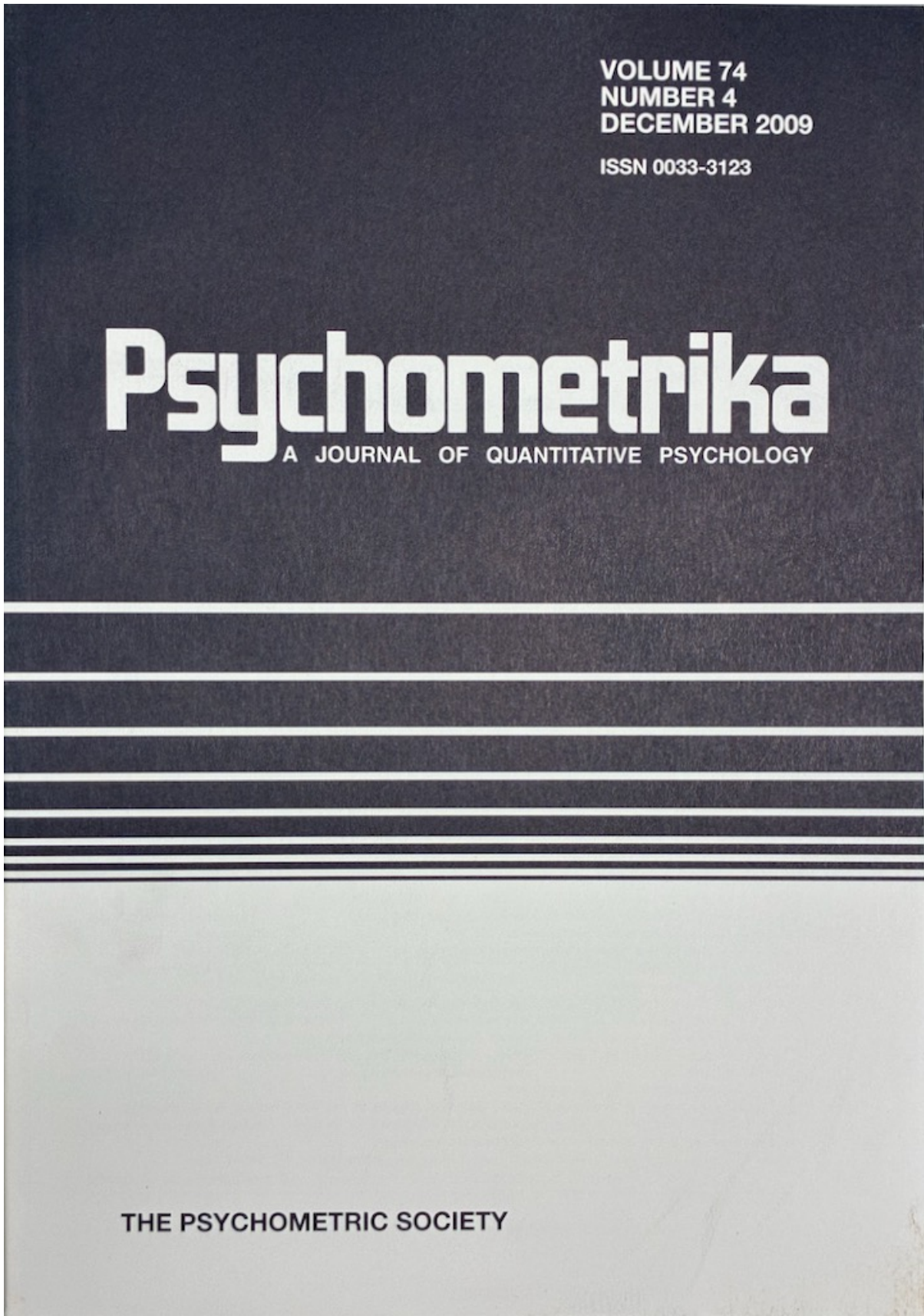


<https://www.psychometricsociety.org/history>

Psychometrika Cover Art

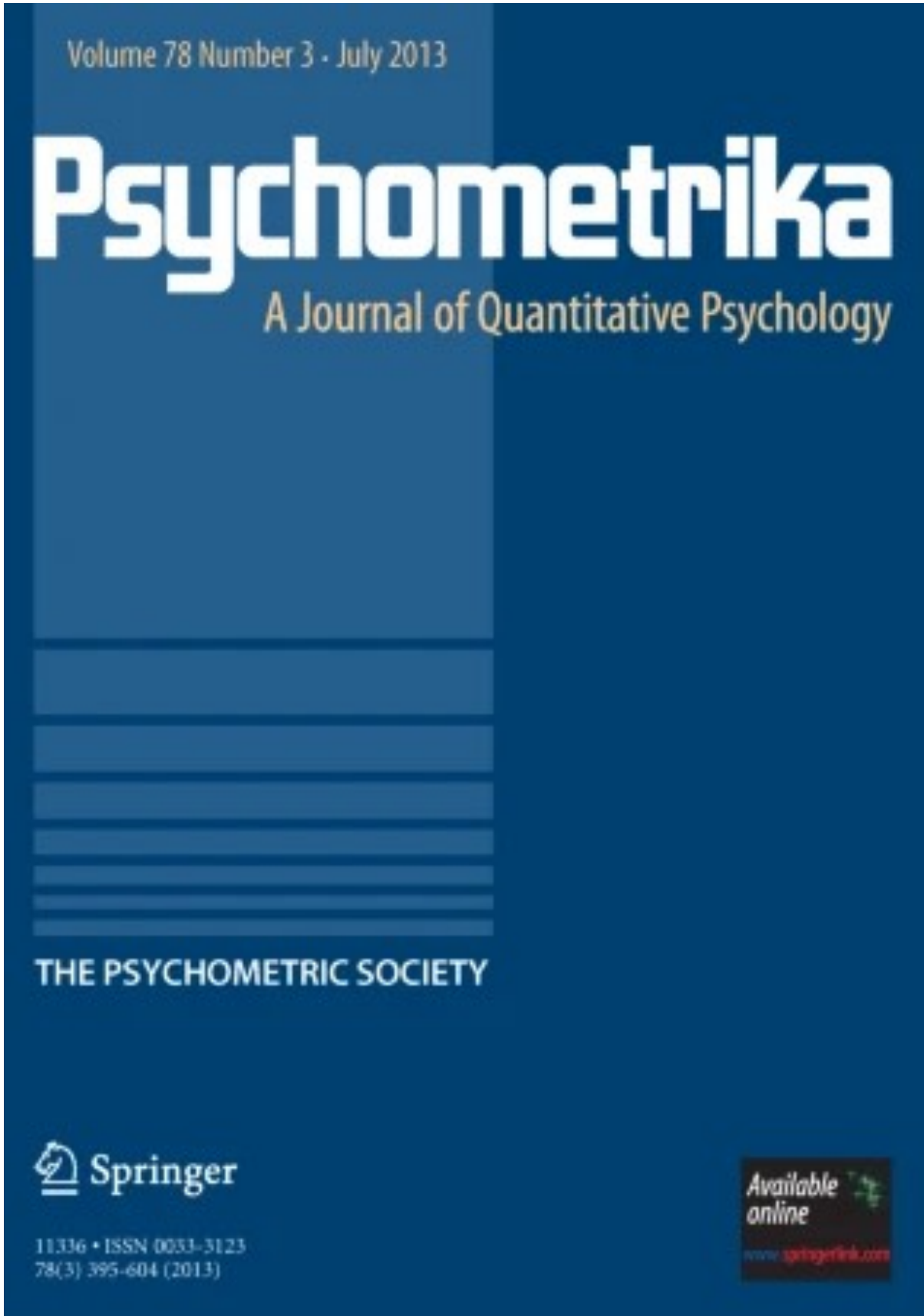


1935-1983

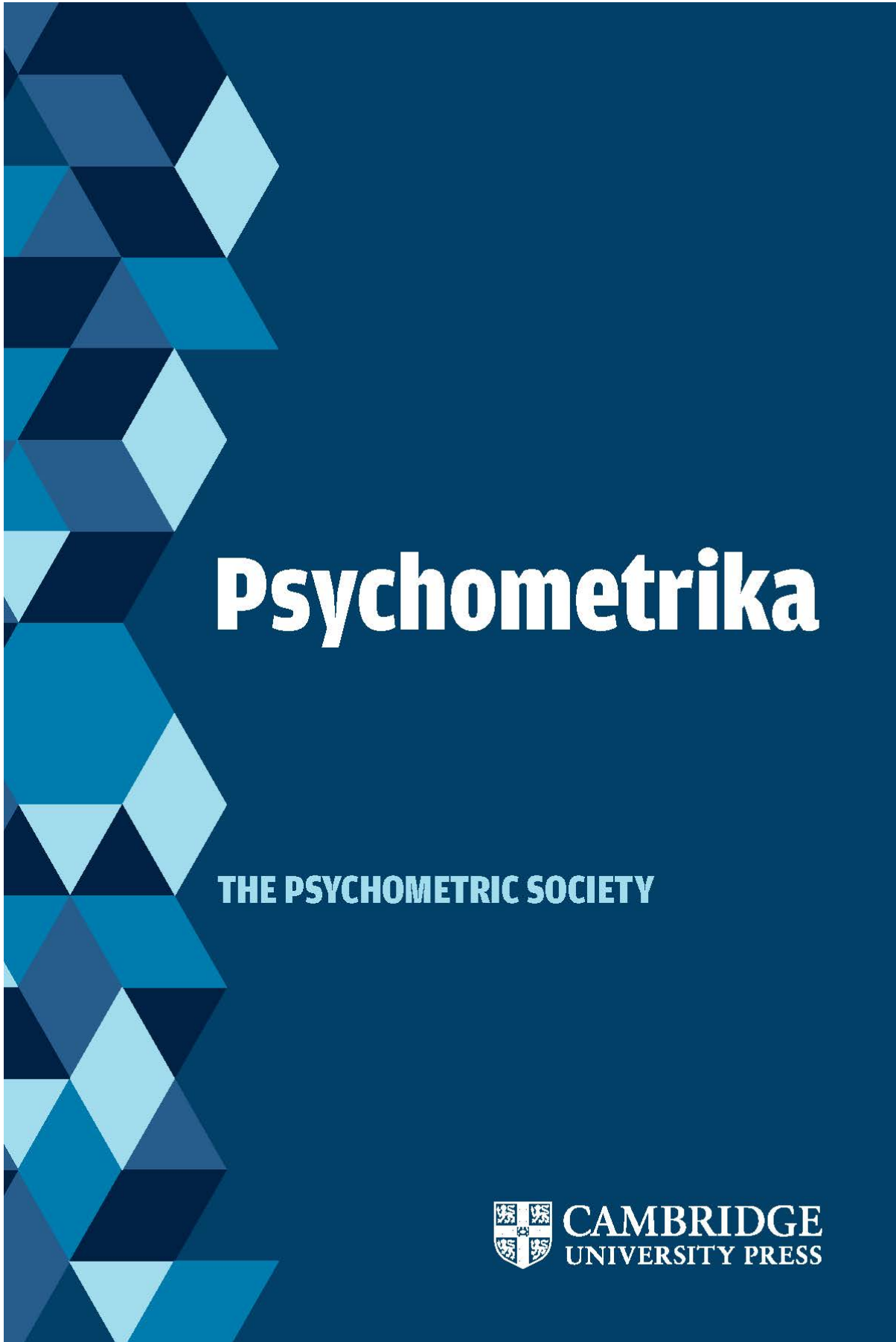


1984-2009

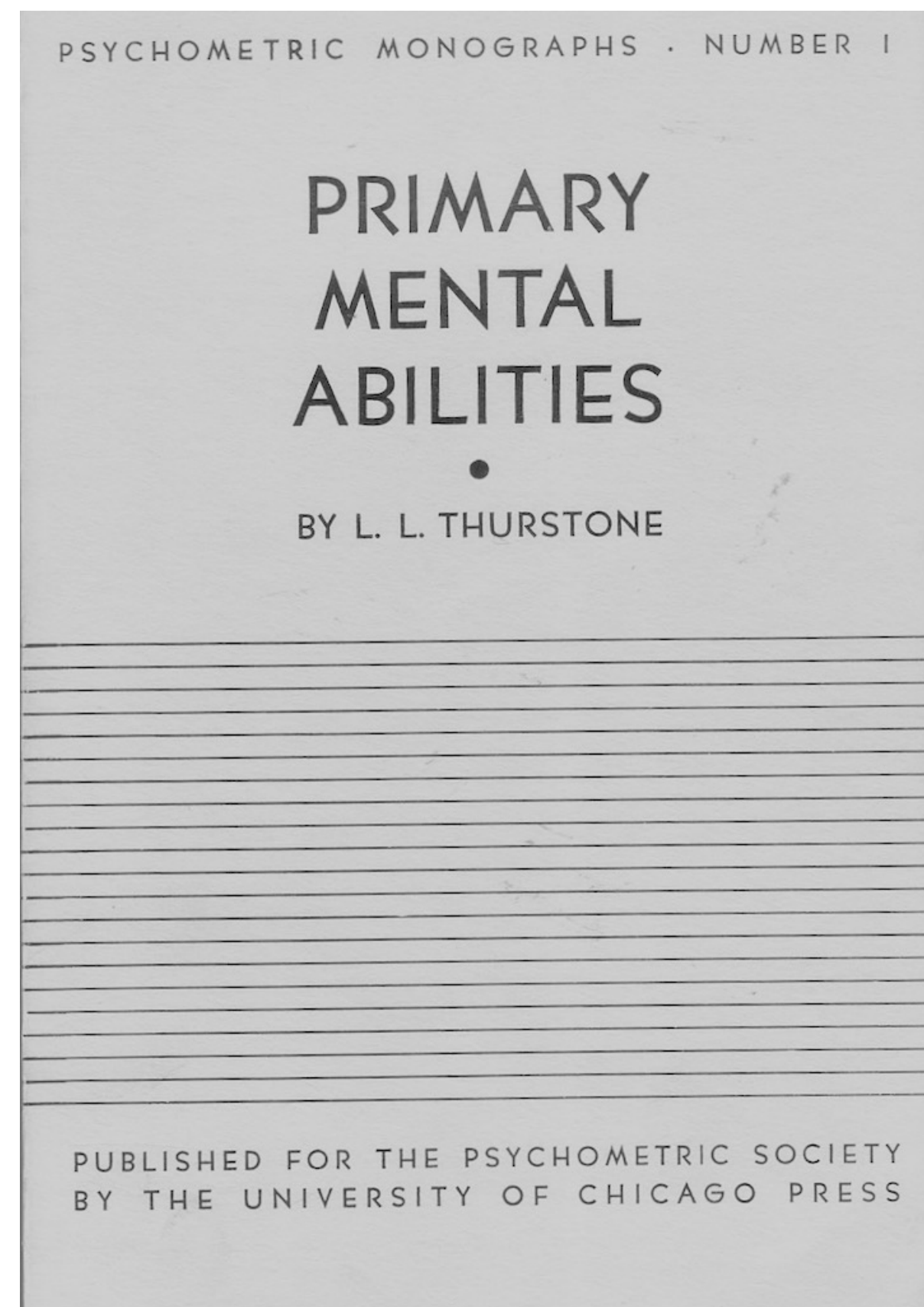
2010-2024



2025-

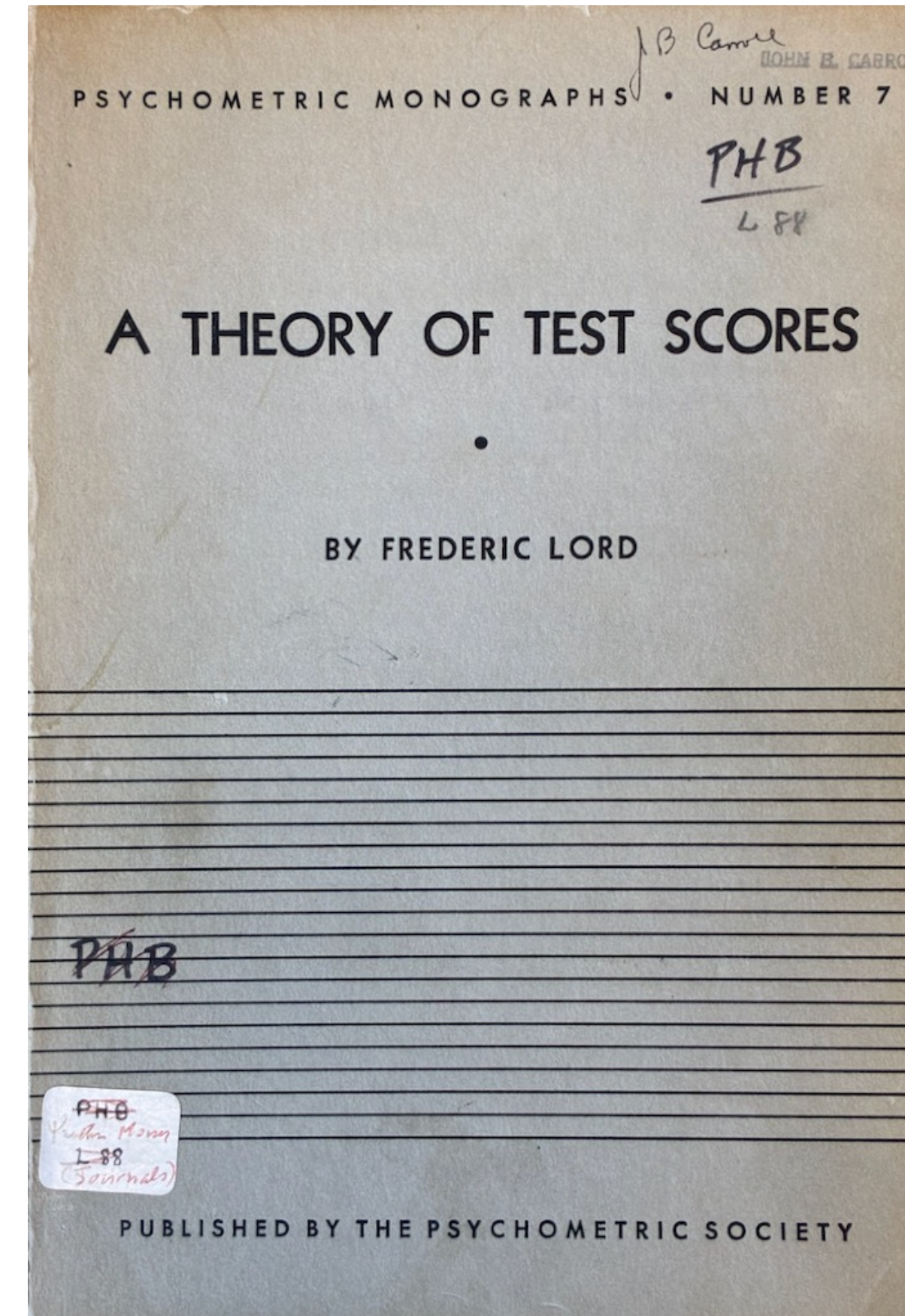
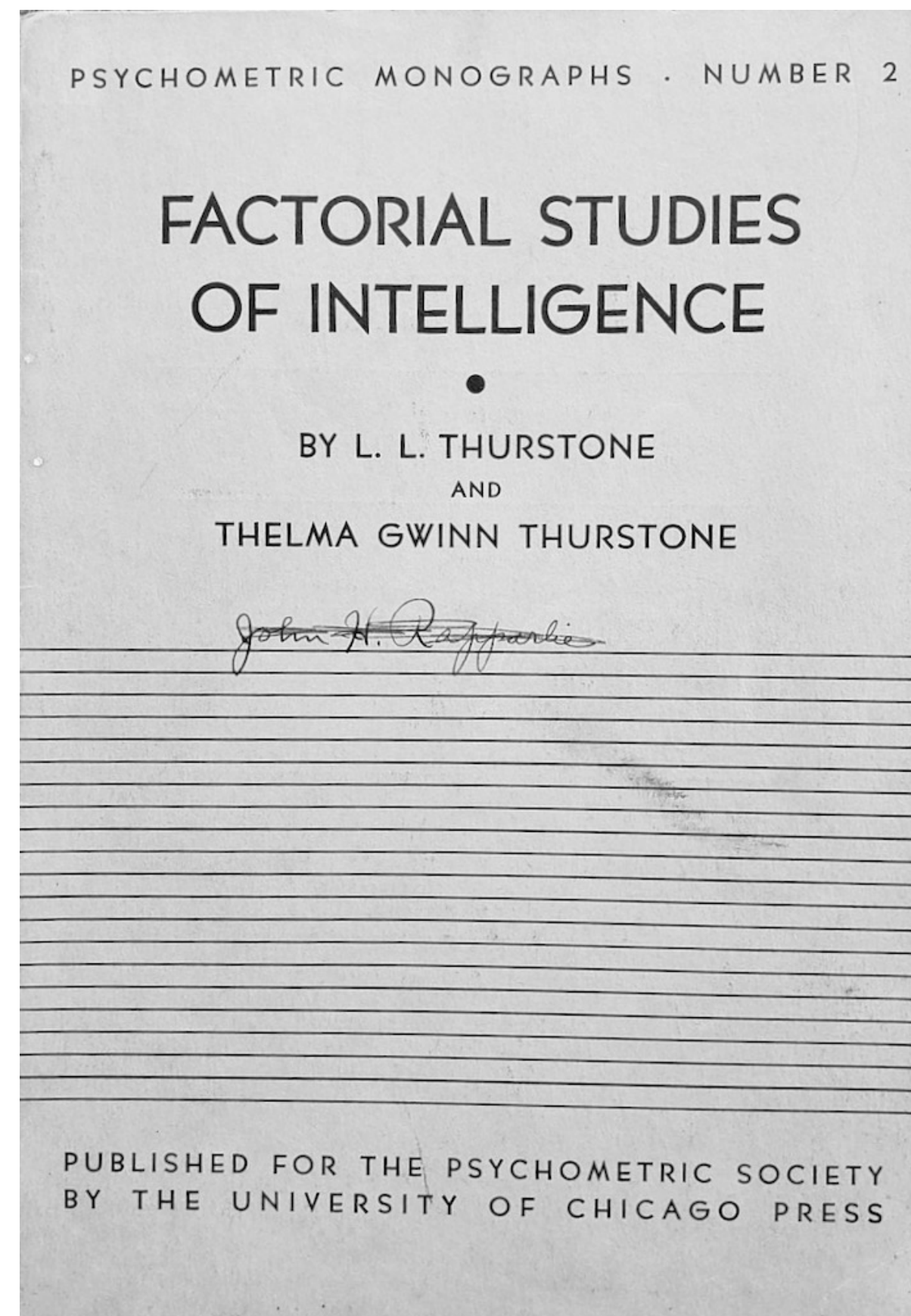


Psychometric Monographs



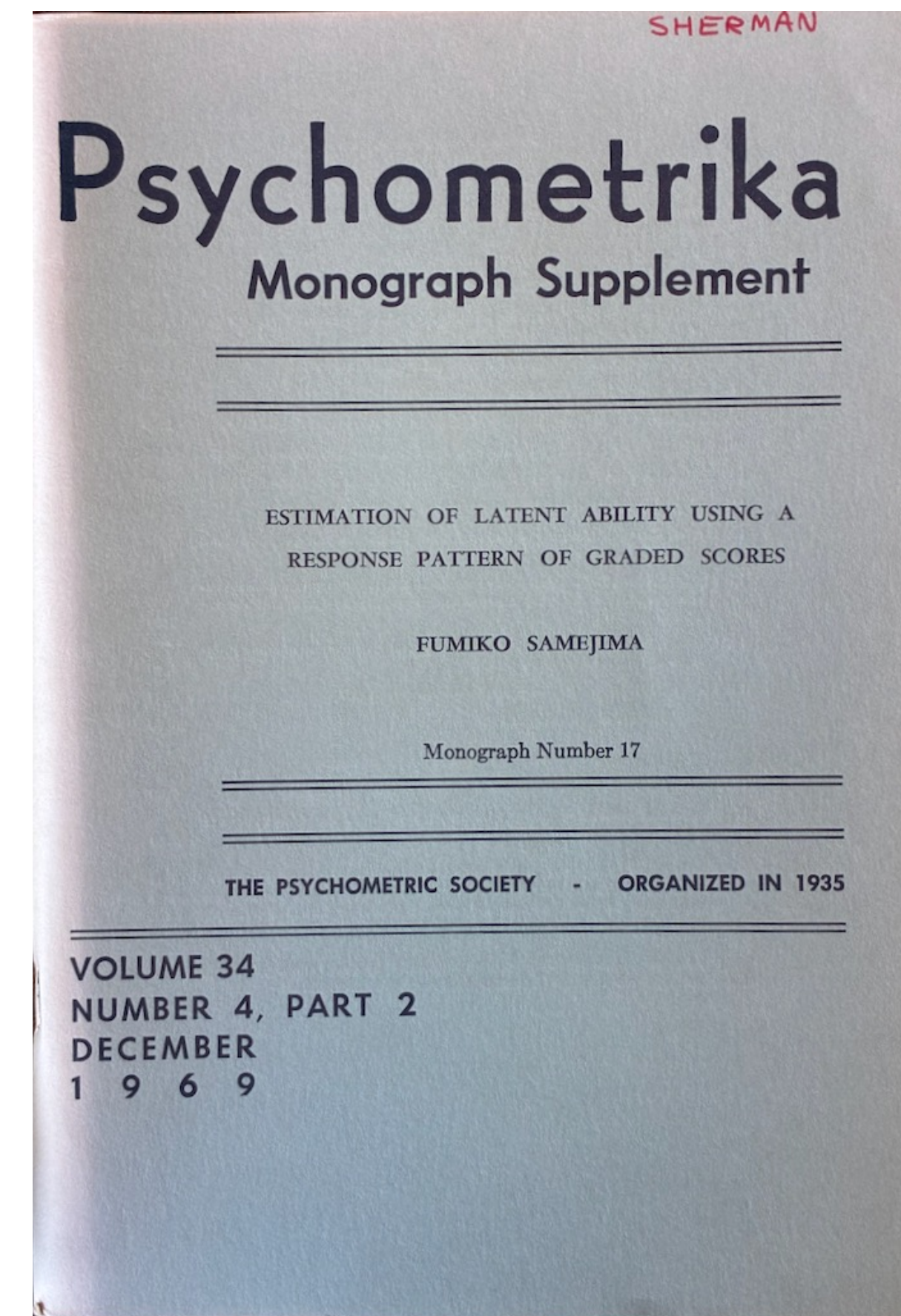
Thurstone, 1938

Thurstone
& Thurstone, 1941



Lord, 1952

Samejima, 1968



The 5 most seminal *Psychometrika* papers

(and some additional thoughts; with a little help from my friend Brian Junker)

Klaas Sijtsma



Contribution to Discussion on Future of *Psychometrika*

IMPS2026, Yonsei University, Seoul

The five most seminal papers in *Psychometrika* (based on # citations WoS, reported on PM's website)

1. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
2. Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
3. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.
4. Tucker, L. R, & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
5. Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

What stands out?

- Top 5 papers (1951-1974) are about reliability (repeatability) and factor structure (dimensionality) of measurement; classical topics
- Next 15 papers have rather diverse topics: multidimensional scaling (3) cluster analysis (2), goodness-of-fit (2; including model selection), item response theory (1), but also factor analysis (4)
- Topics are traditional in the sense of long standing and central to psychological research and research in other disciplines
- Today, Psychometrika's content has become **more diverse** but also **less well recognizable** to researchers (and practitioners); come back to later

In 2016, the ten most cited papers published in *Psychometrika* were discussed in a paper that appeared on occasion of *Psychometrika's* 80th anniversary:

Heiser, W., Hubert, L., Kiers, H., Köhn, H.-F., Lewis, C., Meulman, J., McArdle, J., Sijtsma, K., & Takane, Y. (2016). Commentaries on the ten most highly cited Psychometrika articles from 1936 to the present. *Psychometrika*, 81, 1177-1211.

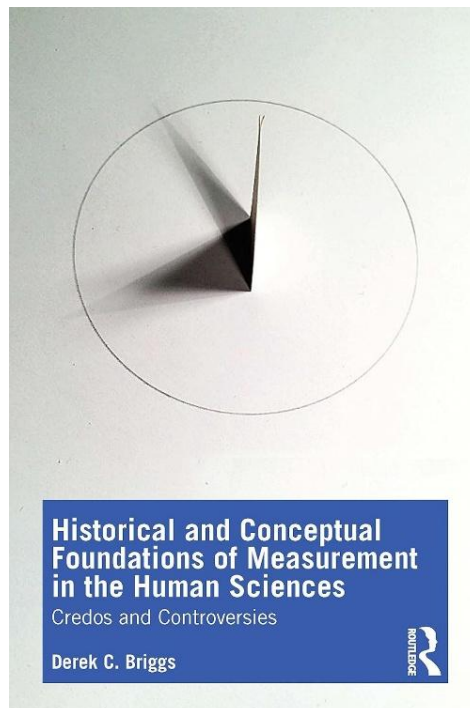
Mission statement on *Psychometrika*'s website

"Psychometrika is a peer-reviewed journal devoted to fostering psychology as a quantitative rational science by examining statistical methods, discussing mathematical techniques and tools useful to researchers and practitioners, and advancing theory for evaluating behavioral data in psychology, education, and the social and behavioral sciences generally."

Conclusion: *Psychometrika* is no longer a journal primarily devoted to **measurement** issues

Question: Does this mean that measurement problems in psychology are **solved**?

Derek Briggs (2021) provides discussion on historical roots of measurement in psychology



- Discusses contributions to psychological measurement by Fechner, Galton, Binet, Spearman, Thurstone, and Stevens
- Recurring theme: nature (processes, structure) of psychological attributes remains unsettled, assumptions of models do not relate to proven characteristics of attributes; thus, **measurement problems unsolved**

Conclusion is also **valid** for more **recent psychometric models**, such as latent class models, item response models, proximity models, and cognitive diagnostic models

They serve as **shallow**¹ models efficiently summarizing data features thus facilitating prediction and decision-making; cf. regression and factor analysis

Deep models explain in their structure the actual physical mechanisms of the phenomenon being modelled

- This kind of research and modelling would **change the kind of papers** *Psychometrika* publishes, and
- Would surely help **improve measurement quality** in psychology and **other human sciences**, but do we want this?

Thanks are due to Brian Junker and Bill Stout, and an untraceable article mentioning the distinction

What is **clear to me** and might be the **focus of discussion today** are the following (discussion) issues

- Psychometrics using **shallow** modeling has been highly **impactful** in **all disciplines** measuring psychological or psychology-related attributes
- **Deep** modeling can help further develop **substantive theory** of (psychological) attributes
- Do we want to include **deep**-modeling papers in *Psychometrika* or are we happy proposing **shallow** modeling (that is of tremendous help summarizing data, no worries)
- I think that **physics** has a fruitful tradition of theory and measurement mutually influencing one another; can we follow their example?

My **conclusion** is that

- Psychometrics is **not done solving** the measurement problem(s) for non-physical attributes; I think **both** shallow and deep modelling can find a place in *Psychometrika* (but perhaps shallow modelling more?)
- Need more attention to **shallow** modelling concerning **test scores**:
 - **Measurement precision**; precision of individual scale scores
 - **Change measurement**; locally precise but general wisdom claims opposite
 - **Person-Fit Assessment**; adds diagnostic information to test score
 - **Test Scores**; the ones practitioners and their clients understand
 - **Validity**; not easily caught in formulas but most important issue of all (shallow: criterion-related validity; **deep**: construct validity)

Thank you for your attention



Where we are now and where we are going

Yunxiao Chen, LSE

Aims and Scope

- *Psychometrika*, the official journal of the Psychometric Society, is devoted to the **development of psychology as a quantitative rational science**. Articles examine statistical methods, discuss mathematical techniques, and advance theory for **evaluating behavioural data** in psychology, education, and the **social and behavioural sciences** generally.
- *The measurement science for social and behavioural research.*

Article Classification (2025-26)

- A (rough) classification of 113 research articles published across five issues in 2025 and three issues in 2026 to date.

Topic	Number (Proportion)	Topic	Number (Proportion)
IRT and CTT	25 (22.1%)	Causal inference	9 (8.0%)
FA, SEM, and latent variable structure	18 (15.9%)	Longitudinal and dynamic models	10 (8.8%)
DCM	11 (9.7%)	Process data, response-time	12 (10.6%)
Missing-data, robust inference, and general estimation	13 (11.5%)	High-dimensional and complex data integration	4 (3.5%)
Clustering, networks, ranking, rater, and choice models	11 (9.7%)		

Strategic Directions

- Within the scope of the journal, the psychometric field should work closely with researchers in social and behavioral sciences to solve new problems in these disciplines.
- AI will become an important topic in these disciplines. For example,
 - AI behavioral science (Jackson et al., 2026) has become a new field in psychology.
 - AI automated item generation and automated scoring.
- With the movement towards "Machine Psychology," Psychometrika needs to expand its scope to include the measurement of non-human agents.
- New measurement tools are emerging to measure complex human traits.

Addressing Complexities

- Many new measurement problems and issues regarding the validity, reliability, and fairness will emerge.
- Addressing these questions requires us to work with more complex data: Process (log) data, ecological momentary assessment data, text data, and network data.

Broadening Submissions

- By expanding our thematic focus (e.g., to AI evaluation), we can attract a broader range of submissions, including those from communities currently underrepresented in the journal.
 - For example, the AI community has increasingly applied psychometric principles and models to AI evaluation—research that currently appears in conference proceedings or Nature-branded journals. We are uniquely positioned to attract submissions from this community.



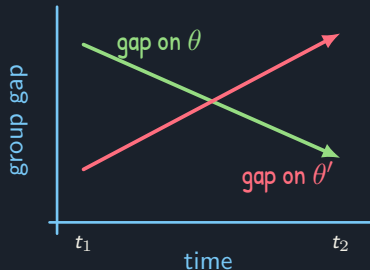
SHOULD WE REPORT ON A LATENT SCALE?

Matthew S. Johnson | Psychometrika at 90 (IMPS 2026)

A provocation.



Reporting on the θ -scale reports a choice, not a fact.

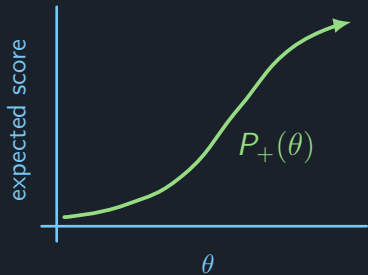


- * “Months of learning” & gap sizes live on a metric we simply *picked*.
- * A gap that shrinks on θ can grow on $\theta' = g(\theta)$: same data, opposite headline. Differences aren't invariant; only order is.

Any monotone g : *same model, same fit*; but the gap's trend can flip.

The unit of θ is simply a convention we chose.

Report what survives $g(\cdot)$.



- * The expected score on a fixed basket of items: an *observable*, invariant to $\theta' = g(\theta)$.
- * **Michell:** we never showed θ is even a *quantity* (double cancellation untested).
- Percentile ranks are preserved under transformation $\pi_i = F(\theta) = F'(\theta')$
- * CDMs / latent class memberships???

$$P_+(\theta) = \sum_j P_j(\theta)$$

The basket is a choice too: an open one, not a hidden rescaling.

But wait, what? Isn't $\theta = F^{-1}(p) = F^{-1}(F'(\theta'))$ preserved under transformations?

Should the 3PL be laid to rest?

No: if we're doing IRT, it's a perfectly good model.

Wainer: "I can always find someone in a ditch with $P = 0$ of getting it right." But is that person in our population?

Guessing is a choice about where structure lives.

- * Put it in the *item*: the 3PL lower asymptote c_j .
- * Or put it in the *population*: a 2PL with a left-truncated θ density is a guessing floor.

The IRF and the population model for θ are coupled: same underdetermination as the scale itself.

Keep the 3PL: its guessing floor is really about the population, not just the item.



motion denied

Psychometrika at 90: Past, Present, and Future

Should the name of “***Psychometrika***” be replaced by
something else that more appropriately reflects the current focus of the journal? 🤔

Psychometrika = Psychological Measurement

- Psyche (ψυχή) ⇒ Psycho-
- metron (μέτρον) ⇒ -metrika (← *Biometrika* by Galton and Pearson)

Recent Publications (selected the first paper from several recent issues)

- 91(3): A Hierarchical Gaussian Process Approach to Understanding Mental Health Trajectories via Item Response Theory Models
- 91(2): A Likelihood-Based Profile Shrinkage Algorithm for Efficient Cognitive Diagnostic Computerized Adaptive Testing
- 91(1): A Recursive Stochastic Algorithm for Real-Time Online Parameter Estimation in Item Response Theory
- 90(5): Detecting Differential Item Functioning across Multiple Groups Using Group Pairwise Penalty
- 90(4): Bayesian Structural Equation Envelope Model
- 90(3): Logistic Multidimensional Data Analysis for Ordinal Response Variables Using a Cumulative Link Function
- 90(2): Show Me Some ID: A Universal Identification Program for Structural Equation Models
- 90(1): Multi-Group Regularized Gaussian Variational Estimation: Fast Detection of DIF
- ...

**Q. Are these not about
Psychological Measurement?**

Psychometrika at 90: Past, Present, and Future

Psychometrika = Psychological Measurement

- Psyche (ψυχή) ⇒ Psycho-
- metron (μέτρον) ⇒ -metrika (← Biometrika by Galton and Pearson)

Recent Publications: Flourishing of Methodological Diversity

- 91(3): A Hierarchical [Gaussian Process](#) Approach to Understanding Mental Health Trajectories via Item Response Theory Models
- 91(2): A Likelihood-Based [Profile Shrinkage](#) Algorithm for Efficient [Cognitive Diagnostic Computerized Adaptive Testing](#)
- 91(1): A Recursive Stochastic Algorithm for [Real-Time Online Parameter Estimation](#) in Item Response Theory
- 90(5): Detecting Differential Item Functioning across Multiple Groups Using [Group Pairwise Penalty](#)
- 90(4): [Bayesian](#) Structural Equation [Envelope](#) Model
- 90(3): [Logistic Multidimensional Data Analysis](#) for Ordinal Response Variables Using a Cumulative Link Function
- 90(2): Show Me Some ID: A Universal [Identification Program](#) for Structural Equation Models
- 90(1): Multi-Group [Regularized Gaussian Variational Estimation](#): Fast Detection of DIF
- ...

⇒ **Growing Diversity in Pursuit of a Common Goal**



Psychometrika, the official journal of the Psychometric Society, is devoted to *the development of psychology as a quantitative rational science*. Articles examine statistical methods, discuss mathematical techniques, and advance theory for evaluating behavioral data in psychology, education, and the social and behavioral sciences generally.

- Theory and Methods
- Application Reviews and Case Studies
- Book Reviews



Yet... we cannot avoid considering the role of AI tools in generating and applying scientific knowledge.

Journals may evolve to prioritize scientific insights beyond what AI can generate

In applied research, when an established methodology exists, identifying the appropriate search terms and leveraging AI tools may substantially reduce the effort. Then... the value of research papers may shift toward demonstrating rigorous scientific judgment.



Methodology focused: Comparably less susceptible to AI

Application focused: data quality, identification of appropriate AI tools & search strategies, evaluation of results and implications



The traditional research paper (e.g., PDF) may remain the primary outlet for scientific communication...

however... 

Educational Measurement: Issues and Practice, Instructional Topics in Educational Measurement Series (ITEMS)

Example: <https://ncme.org/resources/professional-learning/items/module-41-process-data/>

Note on `cc_data$seq`

Creating the `proc object` from our previously preprocessed logfiles (`log_cc_recoded`) is easy.

Requires:

- `action_seqs`: list of action sequences for each examinee (each element is a vector, i.e., the Coding column for that examinee)
- `time_seqs`: list of timestamp sequences for each examinee (i.e., the time column for that examinee)

Note: if you don't have timestamps, can use 1, 2, ..., L_i .

- `ids`: vector of examinee IDs

```
```{r}
log_cc_recoded %>% group_map(# apply to each group, i.e., each student's logfile
 ~ .x$Coding # extract the Coding column for each examinee (list of action seqs)
) -> action_list
log_cc_recoded %>% group_map(
 ~ .x$time # extract the time column for each examinee (list of time seqs)
) -> timestamp_list
log_cc_recoded %>% group_keys() -> Student_IDs # student IDs
```

```
create the proc object:
log_cc_proc <- proc(
 action_seqs = action_list,
 time_seqs = timestamp_list,
 ids = Student_IDs$StudentID
)
```

```
log_cc_proc
```
```

'proc' object of 31559 processes

First 5 processes:

ARE000000100006

| | Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 |
|-------|--------|--------|--------|--------|--------|--------|--------|
| Event | start | reset | 0_0_0 | 0_0_0 | 0_0_0 | 0_0_0 | end |
| Time | 934.1 | 951.0 | 962.3 | 968.6 | 970.4 | 971.0 | 974.5 |

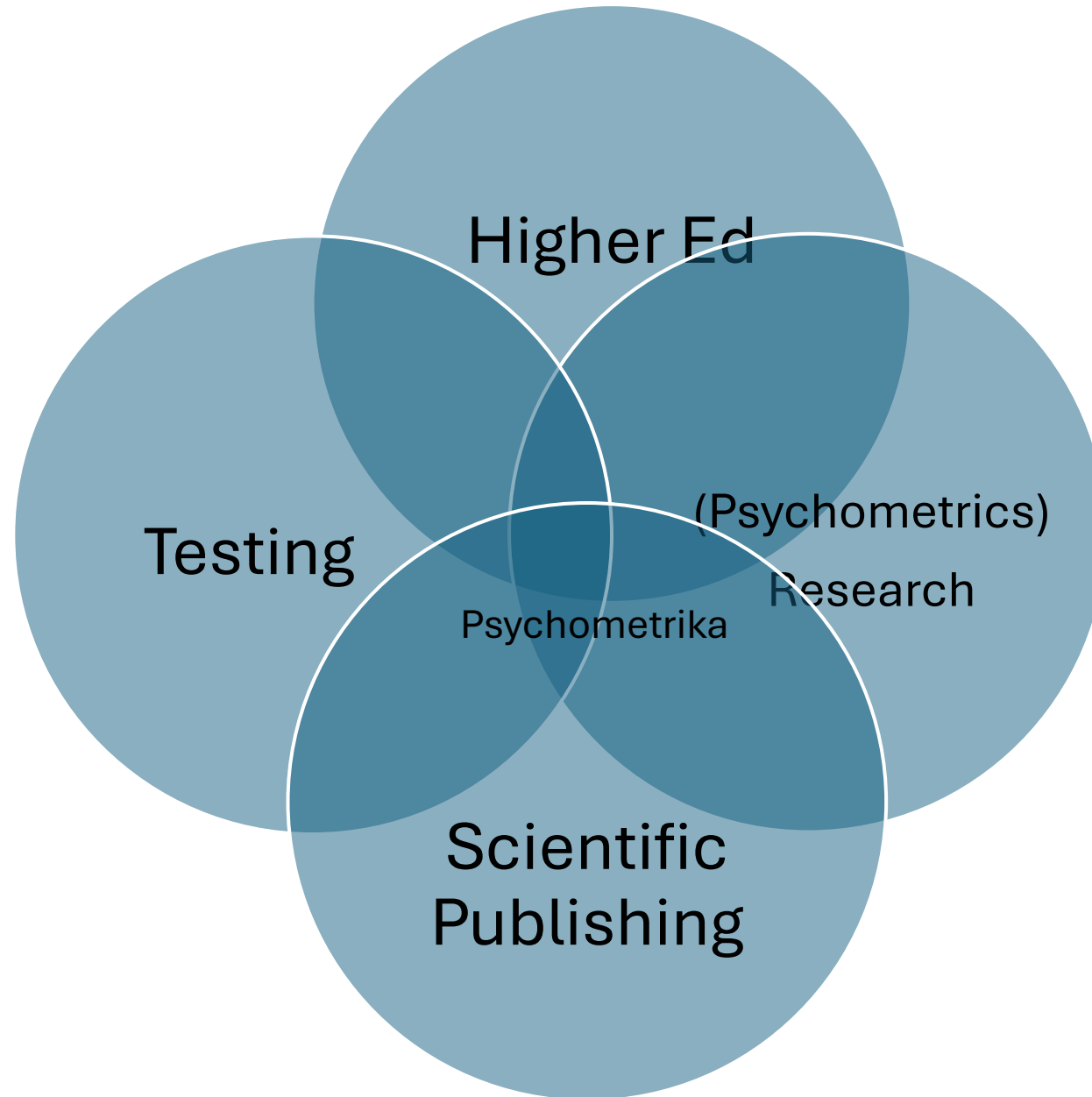
ARE000000100018

| | Step 1 | Step 2 | Step 3 | Step 4 | Step 5 | Step 6 | Step 7 | Step 8 | Step 9 | Step 10 | Step 11 |
|-------|--------|--------|--------|--------|--------|--------|--------|---------|--------|---------|---------|
| Event | start | 1 1 1 | 1 1 0 | 1 1 -1 | 1 1 2 | 0 0 0 | reset | -1 1 -1 | reset | reset | end |

18:33

The Unsaid

- I wonder how, I wonder why...
- Should Psychometrika adapt to AI?
- Why should Psychometrika adapt to AI?
- How should Psychometrika adapt to AI?



How Psychometric Research is Adapting to AI?

- How AI can help move the field of psychometrics forward?
 - Automated item generation (AIG)
 - Personalized testing (CAT), personalized learning
 - Automated scoring
 - Simulated participants/test takers
 - Automated flagging of items with DIF
- How can Psychometrics help move the field of AI forward?
- Research that connects Psychometrics and AI/ML

Potential Growth Areas: The Latter Two

- How can Psychometrics help move the field of AI forward?
 - Li, et al. (ICML, 2026): Use IRT, linking, and CAT to more efficiently benchmark LLM performance
 - Suk & Han (2024): Differential algorithmic functioning
 - Johnson & McCaffrey (2022): Use psychometrics to evaluate algorithmic bias in automated scoring
- Research that connects Psychometrics and AI/ML
 - Cheng (2026): Fairness in AI/ML & psychometrics
 - Suk & Lyu (2026): Causal DIF
 - Gu (2024): Deep CDM

Psychometrika at 90: Past, Present, and Future—Provocations

Daniel Bolt

Dept of Educational Psychology
University of Wisconsin, Madison

July, 2026

What would an ideal peer-review system look like?

Contemporary Challenges:

1. Participation in the review process—broadening participation, increasing rigor
2. Echo chamber—reviewers are often proponents of the same ideas, methods

Increasing Reviewer Participation

1. Publish reviewer reports/commentaries (e.g., “Measurement: Interdisciplinary Research and Perspectives”, under Mark Wilson)—Psychometrika focus articles
2. Incentivizing review process (Pay reviewers? Require reviewing to publish?)
3. Personalizing review requests (clarify why reviewer is being invited)

Addressing Echo Chamber Problem

1. Regularly including “in-” and “out”-of-area reviewers (e.g., editorial board)
2. Use of preprint services that allow community evaluation of manuscripts before formal review