

Test reviewing in Germany

**Carmen Hagemeister, Martin Kersting &
Gerhard Stemmler**

Amsterdam, 4th July 2012



Overview

Source and frame of German test reviewing system

Criteria for test review

German Test Review System

Review process

Review categories

Changes from 2006 to 2009

Difference between German and EFPA/Dutch system

Up to now

At present

What is still necessary

Source and frame of German test reviewing system

DIN 33430

- German standard DIN 33430 for proficiency assessment procedures
- Standard for the process (!)
- Irrespective of profession of assessor
- Passed 2002

- Before DIN 33430:
Long discussions if there should be a test standard.
Decision was: NO!

DIN 33430 contains demands relating to

- Instruments - all instruments, not only tests
- Their application in proficiency assessment procedures

Criteria for test review

Statements extracted from DIN 33430 standards, 140 refer to test manuals

- => can be applied to tests used in any field of assessment

Demands

- information is provided about
 - construction of the test
 - empirical studies
- how the test is
 - applied,
 - scored and
 - interpreted.

Demands for all tests in any field

=> Test reviewing criteria

German Test Review System

Prior to 2006

- List of criteria to use test reviewing, but this list mainly provided definitions of the criteria (Testkuratorium der Föderation deutscher Psychologenverbände, 1986).
- No description of review process
- No ratings intended

Review process

1. Board of Assessment and Testing **chooses test** (anyone can propose tests)
2. Board
 - **mandates 2 reviewers**
 - vouches for **independence and impartiality of the reviewers**criteria of the German research funding organization (DFG) as examples – checked by author
3. Reviewers are provided with **test and necessary material**
Option of “confidentiality”, the Board warrants the confidentiality of information which might be relevant under the law of competition.

Review process

4. Reviewing process:

4.1 Checking if test is eligible for review

DIN Screen Checklist No. 1 (Kersting, 2008): part of the reviewing process, but not published

If not: statement "The test does not meet the demands regarding information and documentation as required in DIN 33430."

=> no further reviewing necessary

This statement is considered as a review.

4.2 Categorizing the test according to the systems of EFPA and ZPID

4.3 Reviewing according to categories of German test review system

Review process

1. 4.3 Reviewing according to categories of German test review system

- Based on information in test manual
- Reviewers can use further information if it is available to the public

EFPA system: inclusion of material which is not public is allowed

Rationale of German review system: DIN 33430: relevant information about test must be available to the "normal" user

- 7 categories plus summarizing final evaluation

Formal evaluation for categories 1, 3, 5, and 6 is given on a four-point scale
"The test fulfils the demands completely / mostly / partly /not at all".

Review categories

	Format	No. words
1. General information about the test, description of the test and its purpose in assessment	Free plus structured	1,000
2. Theoretical basis of test construction	Free text	1,000
3. Objectivity	Free plus structured	1,000
4. Norms	Free text	1,000
5. Reliability	Free plus structured	1,000
6. Validity (including fairness if requested)	Free plus structured	1,000
7. Further quality criteria (susceptibility to failure, non-fakability, and scaling)	Free text	1,000
8. Final evaluation	Free text	2,000

Review process

5. Review steps 4.1 to 4.3 are performed by the reviewers independently.
6. Board receives 2 reviews
Mutual anonymity of reviewers is revoked
Reviewers are asked to create a common version
7. If reviewers cannot agree on common version, relevant differences are presented.
If reviewers do not agree on whether the test is eligible to review or on the formal ratings, Board decides.
8. Board review to authors of the tests
Authors can state their position on evaluation
Board decides whether reviewers are asked to modify their review
Board can modify review if not satisfied with reviewers' modification
9. Reviews published in Report Psychologie and Psychologische Rundschau,
(plus cooperating journals)
10. Authors in alphabetical order unless other arrangements
Each author has the right to remain anonymous

Changes from 2006 to 2009

- No fee
- No more demand that at least one of the reviewers has a licence according to DIN 33430
 - unrealistic
 - excludes experts from other fields
- "Pairs" of reviewers possible: maximum 4 reviewers

Difference between German and EFPA/Dutch system

No fixed criteria for judging size of coefficients

- No system gives reasons for choice of thresholds
Dutch system: Thresholds depend on “importance” of decision => fixed criteria inadequate
- Coefficients should not be as high as possible but contain as much information as possible for the test user
- No wish for artificially high coefficients (in heterogeneous samples)
- No wish for bias in reporting only favourable results

Up to now

- About a dozen reviews printed
- One more being worked on
- 1 year duration of a review

- 2 to 4 new reviewers per review (no teams like for reviews in the Netherlands)
- Slowly moving into fields which are not professional assessment

At present

- Asked for self nomination of reviewers
- All Board members work as senior reviewers
 - More fields
 - More contacts
 - More coordination within Board

What is still necessary

Publishers

- DIN Screen can help to improve tests
- Sometimes information is available but not presented

Journals

- Reviews according to the system, even if
 - Journal chooses reviewer(s)
 - 1 reviewer only
- ⇒ More comparable

Reviewers

- Calls for reviewers in newsletters of psychologists' association
- Should have a PhD, exceptions possible
- Problem: Experts have often cooperated with the test author

Thank you for your attention!