



Universiteit
Leiden

The Netherlands

Design and evaluation of video portfolios : reliability, generalizability, and validity of an authentic performance assessment for teachers

Bakker, M.E.J.

Citation

Bakker, M. E. J. (2008, December 2). *Design and evaluation of video portfolios : reliability, generalizability, and validity of an authentic performance assessment for teachers*. ICLON PhD Dissertation Series. Leiden University Graduate School of Teaching (ICLON). Retrieved from <https://hdl.handle.net/1887/13353>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/13353>

Note: To cite this publication please use the final published version (if applicable).

Design and evaluation of video portfolios

**Reliability, generalizability, and validity of an authentic
performance assessment for teachers**

ICLON

Leiden University Graduate School of Teaching

ico

This research was carried out in the context of the Interuniversity Center for Educational Research



Netherlands Organisation for Scientific Research

This research was funded by the Netherlands Organization for Scientific Research (NWO) (Project no. 411-02-207).

Title: Design and evaluation of video portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teachers

Titel: Design en evaluatie van videodossiers: De betrouwbaarheid, generaliseerbaarheid en validiteit van competentiebeoordelingen bij docenten

Print: Mostert & Van Onderen! Leiden

Cover design: Simon Ouwerkerk (www.ouwerkerk.tv)

ISBN 978-90-804722-8-0

© 2008, Mirjam Bakker

All rights reserved. No part of this thesis may be reproduced, stored in retrieval systems, or transmitted in any form of by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the author.

Design and evaluation of video portfolios

Reliability, generalizability, and validity of an authentic performance assessment for teachers

Proefschrift

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. P.F. van der Heijden,

volgens besluit van het College voor Promoties

te verdedigen op dinsdag 2 december 2008

klokke 13.45 uur

door

Maria Elisabeth Jacomina Bakker

geboren te Beverwijk

in 1977

Promotiecommissie

Promotores

Prof. dr. N. Verloop

Prof. dr. D. Beijaard

Copromotores

Dr. E.C. Roelofs

Dr. E.H. Tigelaar

Referent

Prof. dr. J.M.G. Brekelmans

Overige leden

Prof. dr. P.F. Sanders

Prof. dr. J.H. van Driel

Prof. dr. H. Hulshof

Table of contents

1. Introduction	9
1.1 Background to the study	10
1.1.1 <i>New forms of teacher assessment</i>	10
1.1.2 <i>Validity and reliability issues in authentic performance assessment</i>	11
1.1.3 <i>Validity arguments for authentic performance assessments</i>	12
1.1.4 <i>Measures to reduce the threats to validity</i>	14
1.2 Problem definition and research questions	16
1.3 Context of the study: Assessing teachers' coaching competence	19
1.3.1 <i>Competence-based and self-regulated learning in senior secondary vocational education</i>	19
1.3.2 <i>Competent coaching</i>	19
1.3.3 <i>Scoring procedure</i>	22
1.3.4 <i>Assessor training</i>	22
1.4 Relevance of the study	23
1.5 Outline of the study	24
References	28
2. Video portfolios: The development and practical utility of an authentic teacher assessment procedure	33
Abstract	33
2.1 Introduction	33
2.2 Development of the assessment procedure	37
2.2.1 <i>Defining teaching interventions that can be marked as coaching interventions</i>	38
2.2.2 <i>Defining criteria and performance levels for competent coaching</i>	40
2.2.3 <i>Content of the video portfolio</i>	43
2.2.4 <i>Scoring method for assessing video portfolios</i>	47
2.2.5 <i>Assessor training</i>	50
2.3 Evaluation of the practical utility of the assessment procedure	51
2.3.1 <i>Participants</i>	51
2.3.2 <i>Procedure</i>	51
2.3.3 <i>Instruments</i>	51
2.3.4 <i>Analysis</i>	52
2.4 Results	53
2.4.1 <i>Interrater agreement</i>	53
2.4.2 <i>Interview study</i>	54
2.5 Conclusion and discussion	57
References	61

3. Reliability and generalizability of performance judgments based on a video portfolio 65

Abstract	65
3.1 Introduction	66
3.2 Validity and reliability in scoring, generalization, and extrapolation	68
3.3 Method	72
3.3.1 <i>Design of the performance assessment procedure</i>	72
3.3.2 <i>Measures to achieve reliable and valid scoring</i>	74
3.3.3 <i>Measure to generalize across video episodes</i>	77
3.3.4 <i>Measures to extrapolate to performance outside the assessment context</i>	77
3.3.5 <i>Participants</i>	78
3.3.6 <i>Data collection</i>	78
3.3.7 <i>Analysis: Assessors' scoring</i>	79
3.3.8 <i>Analysis: Generalization across video episodes</i>	81
3.4 Results	82
3.4.1 <i>Assessors' scoring</i>	82
3.4.2 <i>Generalization across video episodes</i>	85
3.5 Conclusion and discussion	87
References	92

4. The impact of construct-irrelevant variance and construct under-representation in assessing teachers' coaching competence 95

Abstract	95
4.1 Introduction	95
4.2 Threats to validity and reliability	98
4.3 Method	101
4.3.1 <i>Design of the assessment procedure</i>	101
4.3.2 <i>Materials</i>	105
4.3.3 <i>Participants</i>	105
4.3.4 <i>Data collection</i>	105
4.3.5 <i>Analysis</i>	106
4.4 Results	110
4.5 Conclusion and discussion	118
References	123

5. General conclusions and discussion	127
5.1 Overview of the study	127
5.2 Conclusions and discussion	131
5.2.1 <i>Reliability of judgments based on a video portfolio</i>	131
5.2.2 <i>Generalizability of judgments based on a video portfolio</i>	135
5.2.3 <i>Validity of judgments based on a video portfolio</i>	137
5.3 Limitations of the study	139
5.4 Suggestions for future research	140
5.5 Implications for assessment practices	142
References	146
Apendices	149
Nederlandse samenvatting	161
Publications	172
Curriculum Vitae	174
Dankwoord	175
PhD dissertation series	177

