The success of the Apgar score demonstrates the astounding power of an appropriate clinical instrument. This down-to-earth book provides practical advice, underpinned by theoretical principles, on developing and evaluating measurement instruments in all fields of medicine. It equips you to choose the most appropriate instrument for specific purposes.

The book covers measurement theories, methods and criteria for evaluating and selecting instruments. It provides methods to assess measurement properties, such as reliability, validity and responsiveness, and to interpret the results. Worked examples and end-of-chapter assignments use real data and well-known instruments to build your skills at implementation and interpretation through hands-on analysis. This is a perfect course book for students and a perfect companion for professionals/researchers in the medical and health sciences who care about the quality and meaning of the measurements they perform.

- Focuses on the methodology of all measurements in medicine
- Provides a solid background in measurement evaluation theory
- Based on feedback from extensive classroom experience
- End-of-chapter assignments give students hands-on experience with real-life cases
- All data sets and solutions are available online
Practical Guides to Biostatistics and Epidemiology

Series advisors
Susan Ellenberg, University of Pennsylvania School of Medicine
Robert C. Elston, Case Western Reserve University School of Medicine
Brian Everitt, Institute for Psychiatry, King’s College London
Frank Harrell, Vanderbilt University Medical Center Tennessee
Jos W.R. Twisk, VU University Medical Center, Amsterdam

This series of short and practical but authoritative books is for biomedical researchers, clinical investigators, public health researchers, epidemiologists, and non-academic and consulting biostatisticians who work with data from biomedical and epidemiological and genetic studies. Some books explore a modern statistical method and its applications, others may focus on a particular disease or condition and the statistical techniques most commonly used in studying it.

The series is for people who use statistics to answer specific research questions. Books will explain the application of techniques, specifically the use of computational tools, and emphasize the interpretation of results, not the underlying mathematical and statistical theory.

Published in the series
Applied Multilevel Analysis, by Jos W.R. Twisk
Secondary Data Sources for Public Health, by Sarah Boslaugh
Survival Analysis for Epidemiologic and Medical Research, by Steve Selvin
Statistical Learning for Biomedical Data, by James D. Malley, Karen G. Malley and Sinisa Pajevic
Measurement in Medicine
A Practical Guide

Henrica C. W. de Vet
Caroline B. Terwee
Lidwine B. Mokkink
Dirk L. Knol

Department of Epidemiology and Biostatistics
EMGO Institute for Health and Care Research
VU University Medical Center, Amsterdam
## Contents

*Preface*  
ix  

1 Introduction  
1.1 Why this textbook on measurement in medicine? 1  
1.2 Clinimetrics versus psychometrics 2  
1.3 Terminology and definitions 2  
1.4 Scope of measurements in medicine 3  
1.5 For whom is this book written? 4  
1.6 Structure of the book 5  
1.7 Examples, data sets, software and assignments 6  

2 Concepts, theories and models, and types of measurements 7  
2.1 Introduction 7  
2.2 Conceptual models 7  
2.3 Characteristics of measurements 10  
2.4 Conceptual framework: reflective and formative models 13  
2.5 Measurement theories 17  
2.6 Summary 26  

3 Development of a measurement instrument 30  
3.1 Introduction 30  
3.2 Definition and elaboration of the construct to be measured 33  
3.3 Choice of measurement method 35  
3.4 Selecting items 37  
3.5 Scores for items 46  
3.6 Scores for scales and indexes 49  
3.7 Pilot-testing 57  
3.8 Summary 60  

© in this web service Cambridge University Press  
www.cambridge.org
<table>
<thead>
<tr>
<th>4</th>
<th>Field-testing: item reduction and data structure</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Examining the item scores</td>
<td>66</td>
</tr>
<tr>
<td>4.3</td>
<td>Importance of the items</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>Examining the dimensionality of the data: factor analysis</td>
<td>71</td>
</tr>
<tr>
<td>4.5</td>
<td>Internal consistency</td>
<td>80</td>
</tr>
<tr>
<td>4.6</td>
<td>Examining the items in a scale with item response theory</td>
<td>84</td>
</tr>
<tr>
<td>4.7</td>
<td>Field-testing as part of a clinical study</td>
<td>91</td>
</tr>
<tr>
<td>4.8</td>
<td>Summary</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>Reliability</td>
<td>96</td>
</tr>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>96</td>
</tr>
<tr>
<td>5.2</td>
<td>Example</td>
<td>98</td>
</tr>
<tr>
<td>5.3</td>
<td>The concept of reliability</td>
<td>98</td>
</tr>
<tr>
<td>5.4</td>
<td>Parameters for continuous variables</td>
<td>103</td>
</tr>
<tr>
<td>5.5</td>
<td>Parameters for categorical variables</td>
<td>115</td>
</tr>
<tr>
<td>5.6</td>
<td>Interpretation of the parameters</td>
<td>120</td>
</tr>
<tr>
<td>5.7</td>
<td>Which parameter to use in which situation?</td>
<td>123</td>
</tr>
<tr>
<td>5.8</td>
<td>Design of simple reliability studies</td>
<td>124</td>
</tr>
<tr>
<td>5.9</td>
<td>Sample size for reliability studies</td>
<td>126</td>
</tr>
<tr>
<td>5.10</td>
<td>Design of reliability studies for more complex situations</td>
<td>128</td>
</tr>
<tr>
<td>5.11</td>
<td>Generalizability and decision studies</td>
<td>131</td>
</tr>
<tr>
<td>5.12</td>
<td>Cronbach's alpha as a reliability parameter</td>
<td>137</td>
</tr>
<tr>
<td>5.13</td>
<td>Reliability parameters and measurement error obtained by item response theory analysis</td>
<td>139</td>
</tr>
<tr>
<td>5.14</td>
<td>Reliability and computer adaptive testing</td>
<td>141</td>
</tr>
<tr>
<td>5.15</td>
<td>Reliability at group level and individual level</td>
<td>142</td>
</tr>
<tr>
<td>5.16</td>
<td>Improving the reliability of measurements</td>
<td>144</td>
</tr>
<tr>
<td>5.17</td>
<td>Summary</td>
<td>145</td>
</tr>
<tr>
<td>6</td>
<td>Validity</td>
<td>150</td>
</tr>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>150</td>
</tr>
<tr>
<td>6.2</td>
<td>The concept of validity</td>
<td>151</td>
</tr>
<tr>
<td>6.3</td>
<td>Content validity (including face validity)</td>
<td>154</td>
</tr>
<tr>
<td>6.4</td>
<td>Criterion validity</td>
<td>159</td>
</tr>
<tr>
<td>6.5</td>
<td>Construct validity</td>
<td>169</td>
</tr>
<tr>
<td>6.6</td>
<td>Validation in context</td>
<td>191</td>
</tr>
<tr>
<td>6.7</td>
<td>Summary</td>
<td>196</td>
</tr>
</tbody>
</table>
# Contents

## 7 Responsiveness

7.1 Introduction 202  
7.2 The concept of responsiveness 203  
7.3 Criterion approach 206  
7.4 Construct approach 211  
7.5 Inappropriate measures of responsiveness 215  
7.6 Other design issues 220  
7.7 Summary 221

## 8 Interpretability

8.1 Introduction 227  
8.2 The concept of interpretability 228  
8.3 Distribution of scores of the instrument 228  
8.4 Interpretation of single scores 235  
8.5 Interpretation of change scores 241  
8.6 Summary 268

## 9 Systematic reviews of measurement properties

9.1 Introduction 275  
9.2 Research question 276  
9.3 Literature search 278  
9.4 Eligibility criteria 282  
9.5 Selection of articles 283  
9.6 Evaluation of the methodological quality of the included studies 284  
9.7 Data extraction 291  
9.8 Content comparison 294  
9.9 Data synthesis: evaluation of the evidence for adequacy of the measurement properties 296  
9.10 Overall conclusions of the systematic review 300  
9.11 Report on a systematic review of measurement properties 302  
9.12 State of affairs 309  
9.13 Comprehensiveness of systematic reviews of measurement properties 310  
9.14 Summary 311

References 315  
Index 328
Preface

Measuring is the cornerstone of medical research and clinical practice. Therefore, the quality of measurement instruments is crucial. This book offers tools to inform the choice of the best measurement instrument for a specific purpose, methods and criteria to support the development of new instruments, and ways to improve measurements and interpretation of their results.

With this book, we hope to show the reader, among other things,

- why it is usually a bad idea to develop a new measurement instrument
- that objective measures are not better than subjective measures
- that Cronbach's alpha has nothing to do with validity
- why valid instruments do not exist and
- how to improve the reliability of measurements

The book is applicable to all medical and health fields and not directed at a specific clinical discipline. We will not provide the reader with lists of the best measurement instruments for paediatrics, cancer, dementia and so on – but rather with methods for evaluating measurement instruments and criteria for choosing the best ones. So, the focus is on the evaluation of instrument measurement properties, and on the interpretation of their scores.

This book is unique in its integration of methods from different disciplines, such as psychometrics, clinimetrics and biostatistics, guiding researchers and clinicians to the most adequate methods to be used for the development and evaluation of measurements in medicine. It combines theory and practice, and provides numerous examples in the text and in the assignments. The assignments are often accompanied with complete data sets, where the reader can really practise the various analyses.
This book is aimed at master’s students, researchers and interested practitioners in the medical and health sciences. Master’s students on courses on measurements in medical and health sciences now finally have a textbook that delivers the content and methods taught in these courses. Researchers always have to choose adequate measurement instruments when designing a study. This book teaches them how to do that in a scientific way. Researchers who need to develop a new measurement instrument will also find adequate methods in this book. And finally, for medical students and clinicians interested in the quality of measurements they make every day and in their sound interpretation, this book gives guidelines for assessing the quality of the medical literature on measurement issues.

We hope that this book raises interest in and improves the quality of measurements in medicine. We also hope you all enjoy the book and like the examples and assignments. We appreciate feedback on this first edition and welcome suggestions for improvement.

The authors
December 2010
In medicine measurement underpins most clinical decisions. Outcome measures for rheumatoid arthritis clinical trials (OMERACT) is an informal collaborative group of professionals dedicated to improving outcome measurement in the rheumatic disease. The methodologic hallmark of the OMERACT process is captured in the OMERACT filter--truth, discrimination, and feasibility. Using the key elements of the OMERACT filter a comprehensive checklist for evaluating reported measures is provided.