

Book Reviews / Comptes rendus

Yorke, Mantz (2007). *Grading Student Achievement in Higher Education: Signals and Shortcomings*. New York and London: Routledge. Pages: 256. Price: 160.00 USD (hardcover).

Reviewed by Don A. Klinger, Associate Professor, Faculty of Education, Queen's University

One of the more contentious issues in the field of assessment and evaluation is that of grading. Issues of grading are of even greater concern when relatively high-stakes decisions are made based on students' grades. Mantz Yorke's (2007) book, *Grading Student Achievement in Higher Education: Signals and Shortcomings*, plunges into the grading issue in a high-stakes environment: higher education. Certainly, the grades students obtain in institutes of higher education impact their future career and academic pursuits. Yorke examines the issue from a variety of contexts, examining the complexity of grading and the impacts and implications of current grading practices.

With nine chapters and 208 pages of text, the book is a relatively short read. Nonetheless, Yorke addresses a broad range of issues within this text. The first chapter, "The Complexity of Assessment," provides a context for the book. Yorke uses an easy to read style to highlight the major issues in assessment, including the purposes of assessment, formative and summative assessment, norm- and criterion-referenced assessment, reliability and validity. Chapter 2, "Grading and its Limitations," examines the topic of grading from a general perspective. Yorke makes good use of quotes from researchers to highlight the longstanding concerns surrounding grading. Yorke summarizes the scales used in grading, scale size, the links between scales and purpose, grading problems, reliability and consistency. This is the longest chapter in the book but it is a valuable chapter for those responsible for grading, whether in higher education or in other educational contexts.

Chapter 3, "Variations in Assessment Regulations," examines specific issues in grading, using three different countries: Australia, the United Kingdom and the United States. Although Yorke does not provide a real rationale for these three choices, his comments throughout the book suggest that the UK (honours degree classification) and the US (grade-points) were chosen because of their contrasting methods of grading in higher education. Australia was chosen because higher education institutes in Australia tend to use a grading system that

is a hybrid of the other two, albeit with its own unique features. This chapter does highlight the bias that occurs throughout the book, having a greater focus on these three countries. Other grading systems receive only passing mention or are not included. Higher education in Canada is not considered at any point in the text. Nonetheless, the three countries more thoroughly examined do provide contexts that should be familiar for those who work in higher education institutes in other jurisdictions. Yorke makes good use of existing research to highlight the issues of grading in the three countries and the variations that occur across differing institutions in these countries.

As the name suggests, Chapter 4, "UK Honours Degree Classification, 1994-95 to 2001-02," uses data from an eight-year period to examine trends in grading in the UK. Given the importance attached to obtaining a "good" honours degree, Yorke uses the data to examine trends in the proportion of honours degrees being awarded across institutions and programs. As with all of the chapters, this chapter is an engaging read, examining issues of data quality, meaningfulness of findings and trends. Once again, Yorke reiterates the complexity of grading and our attempts to examine it. Underlying the focus of Chapter 4 is the issue of grade inflation. Chapter 5, "How Real is Grade Inflation," delves into this topic in more detail. Yorke acknowledges the trends for increased grades over time and the concerns that exist about grade inflation. Nonetheless, he provides a balanced examination of both aspects of grade inflation and non-inflationary causes of increased grades in higher education. Concerns about grade inflation are real in higher education and this chapter goes far to summarize the major aspects of the issue although it is short on solutions.

Chapter 6, "The Cumulation of Grades," examines in more detail the issues that occur when individual grades are combined to create a single overall grade. The issues in this chapter are technical in nature; however, Yorke is able to present the material in a way that even non-measurement specialists would grasp the fundamental issues. As the chapter illustrates, cumulation of grades is problematic given the variability in individual course grading methods and differences in the determination of cut-scores for differing grades. Cumulation is examined using both the system of Honours Degree Classification and GPA. Yorke acknowledges the potential superiority of GPA but also rightly finds it wanting. Chapter 7, "Value Added," examines the increasing attempts to measure real gain in education. As Yorke acknowledges, value-added models are generally associated with accountability frameworks, and proponents consider them to be more fair as measures of student intake are included in the model. Starting from the widely held measurement perspective that "measuring gain is not simple," the chapter examines attempts to use value-added models. Examples are provided from Australia, the UK and the US, mostly within the contexts of public education. As Yorke concludes, given the variation in determining what value added actually means and the problems of measuring gain in public education, the goal of determining value added in higher education will remain elusive for some time.

“Fuzziness in Assessment” (Chapter 8) approaches the notions of measurement error from a qualitative perspective. Precision in assessment and grading, Yorke argues, is not possible in the context of higher education. Rather he attempts to find out what models would be “good enough” for assessment and grading in higher education. While the chapter does identify issues of precision in assessment, it does not acknowledge the current knowledge and ongoing research designed to increase measurement precision. Rather than simply arguing for a “satisficing” perspective, Yorke needs to place more responsibility on professors to develop their own assessment skills. For example, increasing the assessment skills of professors would result in higher quality assessments having greater precision and far less fuzziness.

Chapter 9, “Judgement, Rather than Measurement,” begins with a critique of the learning outcomes approach within the context of higher education. Yorke offers a “top down” approach as one solution, in which a student’s overall grade is based on their body of work within their program. Such a model is similar to those used in portfolio assessments and Yorke makes the connection. Nonetheless, such alternative assessments have the same technical issues as found in traditional methods, and in many cases they are more severe. The chapter continues, exploring current policy recommendations for grading in higher education, and the needs of employers. As Yorke admits, “Assessment is challenging.” Nonetheless, if higher education is to promote learning approaches in students that move beyond a performance approach focused on “getting the grades,” the grading methods need to also encourage learning approaches. As Yorke concludes, this will only occur with increased discussion on the topic and increased expertise in assessment.

Overall, the book *Grading Student Achievement in Higher Education* provides a more than satisfying summary of the issues in grading, not only for higher education but also for education in general. Throughout the book, Yorke uses relevant literature and analogies to clarify the issues. The text is well written and easy to read. It is engaging and thought provoking. The Australia, UK, and US focus is somewhat limiting but not overly so. Certainly readers from other jurisdictions would find much that fits their own contexts. The book will find an audience in researchers, administrators and policy makers in higher education. There is also much in the book that would satisfy the same audiences within public education. Yorke acknowledges and explores the complexity surrounding the major aspects of assessment and subsequent grading. A critical reader may be a little disappointed in the lack of solutions Yorke provides to addressing these complex issues. There may also be some argument surrounding the central argument linking grading in higher education and employment. Certainly the two are related but Yorke implies the central purpose of higher education is to secure advantageous career opportunities and that grades may either support or hinder this pursuit.

From a measurement perspective, Yorke demonstrates a sound knowledge of both the technical and conceptual aspects of assessment and grading. A

discussion of decision consistency (see for example Cohen, 1960; Kane & Brennan, 1980; Livingston, 1972) would have been useful alongside the section on reliability. Given the nature of the assessment and grading in higher education, measures of decision consistency are likely more important than traditional measures of reliability. An ongoing issue throughout the text is that Yorke blurs the distinctions between assessment and grading. Certainly the two are closely linked as grading decisions require assessment information. Assessment is a measurement concept. In contrast, measurement specialists have long recognized that grading is a judgmental process. The quotes Yorke uses to differentiate measurement and judgement in Chapter 9 recognize this distinction. However, Yorke does not. The solutions to resolve assessment issues are largely, albeit not exclusively, technical in nature. In contrast, the judgemental process of grading will require a different perspective to resolve its ongoing issues. Regardless, as evidenced in Chapter 9, Yorke correctly addresses the judgemental nature of grading.

Academics and policy makers who have an inherent faith in the grading procedures we use in education will do well to read Yorke's book. Those with a critical perspective will certainly find support for their views from research Yorke has cited and the conceptual issues described in the book. Few current grading practices are safe from Yorke's critiques, and his critiques are accurate and telling. Calls for reforms in grading practices in public education have been increasing for more than a decade (see for example Marzano, 2006). Interestingly, the solution Yorke proposes for higher education, using multiple indices, is very similar to that proposed by Marzano. Yorke has made a strong case for the reformation of grading in higher education. Given the nature of higher education institutions, such changes will be difficult to enact. Nevertheless it is time to address these issues. ♣

REFERENCES

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Kane, M. T. & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement*, 4, 105-126.
- Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. *Journal of Educational Measurement*, 9, 13-26.
- Marzano, R. J. (2006). *Classroom assessment and grading that works*. Association for Supervision and Curriculum Development.