Peer-Assessment in Higher Education: A Review of Recent Studies

Michael Mogessie Ashenafi

Department of Information Science and Engineering
University of Trento

06 November, 2015

Introduction

- Assessment in education - varies with goals
- Continuous vs one-off
- Goal - measuring performance and/or improving student learning
- Terminologies - Summative and Formative

Summative Assessment

- intended to measure degree of achievement
- either one-off or carried out at intervals - Mid-terms, finals
- Criterion-referenced (absolute grading) or normative (relative to other students)

Formative Assessment

- student-centered
- Goal - to provide support and feedback to students
- Helps students monitor their own progress
- Also helps the teacher to adjust their instruction accordingly
- Should not contribute towards final grades

Non-Traditional forms of Assessment

- The teacher is not the sole assessor
- significant involvement of students
- purely formative, or a blend
- E.g. Self-assessment, peer-assessment

Peer-Assessment

"... an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status."

Topping(1998)

In this study ...

- Over two-decades of research in peer-assessment
- The million dollar question is - **Does it really work?**
- This review examines recent literature:
  - to find out if there's a clear-cut answer
  - to identify challenges and opportunities
  - to recommend ways to tackle challenges in the practice

Outline

1. Introduction
2. 20th Century Peer-Assessment
3. 21st Century Peer-Assessment
   - Inclusion Factors
   - Themes of Interest
     - Literature Reviews
     - Case studies, action research and peer assessment instruments
4. Discussion
5. Recommendations
6. End of Talk

Topping (1998) - A qualitative study

- Reviewed 109 studies to find out if PA works
- Identified many variables among the studies
  - what subject?
  - nature of the PA task assessed: educational vs. professional
  - formative or summative?
  - what is being assessed?
  - do peer-assigned scores agree with those of the teacher's?
- His conclusion:
  - too many variables
  - no concrete evidence regarding the soundness or practicality of PA in higher education

Falchikov and Goldfinch (2000) - A meta-analytic study

- conducted a meta-analytic review of 56 studies comparing peer and teacher marks
- Variables identified
  - population characteristics
  - work being assessed
  - course level
  - nature of assessment criteria
  - number of teachers and students involved per assessment task
- Their conclusion: On average, student marks agreed with teacher marks:
  - mean *r=0.69* - the higher the better
  - mean effect size *d=0.24* - the lower the better

Falchikov and Goldfinch (2000) - Six Influential Factors

- assessing individual dimensions vs overall judgements using well-specified criteria
- The nature of the assessment task - educational product or process vs. professional practice
- Better experimental designs (e.g. sample sizes) $\rightarrow$ better agreement
- Number of students involved per assessment task
- The subject area - less agreements in medical education
- Involving students in the development of assessment criteria $\rightarrow$ better agreement

Outline

1. Introduction
2. 20th Century Peer-Assessment
3. 21st Century Peer-Assessment
   - Inclusion Factors
   - Themes of Interest
     - Literature Reviews
     - Case studies, action research and peer assessment instruments
4. Discussion
5. Recommendations
6. End of Talk

The Selection Process

- Keywords - peer assessment, peer grading, peer evaluation, peer review, peer feedback, peer interaction
- Google Scholar
- Journal articles and conference proceedings published since 2000
- Not computer-based or web-based (Luxton-Reilly (2009) provides a comprehensive review)
- Final list included 64 studies

Two Main Categories

- Literature Reviews
  - Student involvement
  - Variables of peer-assessment
  - Quality of peer-assessment
- Case studies, action research and peer assessment instruments
  - The value of peer-feedback
  - Peer-assessment design strategies
  - Perceptions of students and teachers
  - Psychological and social factors in peer-assessment
  - Validity and reliability of peer-assessment

Student Involvement

- Several studies recommend that students be actively involved at various stages of PA
- Falchikov (2003), Leenknecht et al. (2011), Bloxham & West (2004), Sluijimans et al. (2004)
  - PA must actively involve students to be effective
  - PA experiments should allow replication
  - clear instructions for students regarding processes involved

Variables of peer-assessment

- Van Zundert et al. (2010) reviewed 26 articles between 1990 and 2007
- Identified four variable categories
  - Psychometric qualities
  - Domain-specific skills
  - Peer-assessment skills
  - students' attitudes towards PA

Variables of peer-assessment

- Topping (2010) - reveals many uncertainties in PA and identifies 17 variables
  - Do peer-peer relationships affect the practice?
  - Should peer-feedback be iterative or one-off?
  - Is assigning multiple students to the same assessment task effective?
  - inconsistencies, contradictory results, flaws or limitations of studies are revealed

Variables of peer-assessment

- Van den berg et al. (2006a) select 10 of Topping's 17 variables
- Important for optimal peer-assessment design
  - What is being assessed? Written work? Oral presentation?, ...
  - Is PA as substitute for teacher's assessment?
  - Is it mutual, anonymous?
  - Is contact face-to-face?
  - in-class, take-home?
  - Are there any incentives?

Variables of peer-assessment

- Van den berg et al. (2006b) build upon previous research
- Impact of variables on oral and written feedback
- Peer-feedback is optimal when:
  - PA conducted in small groups, formative or summative
  - Written feedback should be orally explained and discussed with the assessed
  - But what about large classes?

Quality of Peer-Assessment

- Tillema et al. (2011) - How to measure quality of PA practices
- 3 quality criteria should be met at all stages of the assessment process
  - Authenticity - process needs to actively engage students - representativeness, meaningfulness, cognitive complexity, content coverage
  - Transparency - tasks should be clear, understandable, and doable
  - Generalisability - can outcome be generalised to those of tasks measurin the same achievement? - comparability, reproducibility, educational consequences

1. Introduction
2. 20th Century Peer-Assessment
3. 21st Century Peer-Assessment
   - Inclusion Factors
   - Themes of Interest
     - Literature Reviews
     - Case studies, action research and peer assessment instruments
4. Discussion
5. Recommendations
6. End of Talk

The Value of Peer-Feedback

- Miller (2003) - Quality of peer-feedback determined by specificity of criteria
  - The more specific the criteria, the more discriminative PA is - risks lowering feedback quality
- Strijbos et al. (2010) - Is elaborate feedback good?
  - The majority of 89 grad students didn't think so
  - Adequate but had a negative impact
  - Degree of specificity and brevity have varying impacts on students with different competence levels
- Lin et al. (2001) - In general, specific feedback more helpful than holistic feedback in improving performance

The Value of Peer-Feedback

- Althauser & Darnall (2001), Tsai et al. (2002) - Students who provide high-quality feedback tend to incorporate feedback from peers in their revisions.
- Li et al. (2010) - Strong positive relationship between a student's quality of feedback and the quality of their own final project.
- Cho and McArthur (2010) - Feedback from multiple peers is more helpful than that from just one.
- Hu (2005), Min (2006), Sluijsmans and Prins (2006), Saito (2008) - Training students in providing feedback and in PA skills, in general, improved quality of feedback and work being assessed.
- Chen & Tsai (2009) - Subsequent feedback tends to produce marginal improvement in the quality of work being assessed

Peer-Assessment Design Strategies

- Topping et al. (2000)
- PA conducted in a class of 12 grad students
- Formative
- Product assessed - end-of-second-term academic report
- Mandatory participation, PA results did not contribute to final marks
- Out-of-class, anonymous, reciprocal
- 14 specific criteria provided
- Study sought to investigate peer and teacher score agreements
- Conclusions:
  - Adequate reliability and validity of the approach
  - May, however, not generalise to other settings

Peer-Assessment Design Strategies

- Ballantyne et al. (2002) - One of the largest PA studies
- A three-phase study spanning a two-year period
- 1654 students and 30 staff from three departments
- PA procedures outlined and revised together with students
- Shortcomings - assessment was manual, anonymity was not preserved in some departments
- Increase in student load - required to meet outside class to exchange assignments and agree on final grades, risk of bias
- Otherwise a thoroughly designed high quality study

Peer-Assessment Design Strategies

- Automating peer-assessment tasks has several advantages
- teachers can enjoy PA advantages less the negative impacts discussed
- anonymity, efficient assignment distribution, discussion, and submission of grades easily guaranteed
- automation could also help calibrate grades assigned by multiple peers (Hamer et al. 2005)

Peer-Assessment Design Strategies

- Some variations
  - the teacher assessing the quality of feedback instead of analysing peer-assigned marks (Davies 2006)
  - PA without explicit assessment criteria (Jones & Alcock 2014)

Perceptions of Students and Teachers

- Overall positive perceptions of students reported by:
  - McLaughlin & Simpson (2004), Saito & Fujita (2004), Wen & Tsai (2006), Wen et al. (2008), McGarr & Clifford (2013)
  - Chang (2006), Kwok (2008), Wood & Kruzel (2008), Xlao & Lucking (2008)
- PA is productive and gives me a clearer view of how teachers assess students (Hanrahan & Isaacs 2001)
- Increased responsibility for others and improved learning (Papinczak et al. 2007)
- Time-intensive, intellectually challenging, creates a socially uncomfortable environment (Topping et al. 2000, Hanrahan & Isaacs 2001, Arnold et al. 2005, Praver et al. 2011)

Psychological and Social Factors in Peer-Assessment

- Gender effects are the least studies factors in PA in higher education (Falchikov & Goldfinch 2000, Falchikov 2003, Topping 2010)
- Bias may not be an issue when PA is anonymous
- The most affected are those which involve visual contact between peers
- A study involving 41 undergrads (20 females) found that males rated males slightly higher than female presenters (Langan et al. 2005)
- This was not the case for females - (Langan et al. 2005, Langan et al. 2008)
- A study of 40 students involved in a PA task (20 females) reported that female students found it a stressful task (Pope 2005).

Validity and Reliability of Peer-Assessmentl

- These are the most common studies
- Validity - how similar are teacher and peer marks?
- Reliability - How close are scores assigned by peers (teachers) to the same work? AKA - Inter-rater reliability
- 15 studies were examined
- 8 reported correlation coefficients
- 4 reported mean and standard deviation - effect sizes (d) were computed
- $d = \frac{2*[mean(eg)-mean(cg)]}{sd(eg)+sd(cg)}$
- eg = experimental group, cg = control group

Validity and Reliability of Peer-Assessmentl - Results

- Mean correlation coefficient(r) of **0.80** and mean effect sizes(d) of **0.27**
- Corroborates findings by Falchikov & Goldfinch (2000), although with much smaller studies
- Most studies varied in the design of assessment tasks
  - Products assessed - written work, oral presentation
  - Disciplines - education, business, law, medical education, computer science and engineering
  - Stats reported - correlation coefficients, one-way & multiple ANOVA, Cronbach's alpha, t-tests, intraclass correlation, mean and SD

In Summary

- Focus of this study was on PA in higher education
- Variables of interest have led to a multitude of design strategies
- Commendable studies providing insight into the intricacies of PA practice
  - Cho et al. (2006)
  - Ozogul & Sullivan (2009)
  - Smith et al. (2002)
  - Xiao & Lucking (2008)
  - Sahin (2008)
- Maintaining anonymity in manual PA becomes a luxury as the number of students involved increases
- Lack of common standards - most studies are not readily comparable

In Summary

- Most studies mix experiments and attempt to measure several variables - mixed results?
- No attempts to take advantage of advances in related disciplines
- The vast majority are standalone practices in conventional classrooms
- Advances in computer science are being applied in almost all social systems
- PA has yet to take advantage of these - So far, web-based PA only

In Summary

- Majority PA practices are one-off experiments - how do we test if it helps long-term learning?
- Having PA practice as part of a curriculum is a risky business - who are the stakeholders?
- Most studies are disconnected and only few build upon previous studies
- Lack of studies regarding impacts of gender, race, anonymity, academic dishonesty
- How about impact of formative peer-assessment on students' performance on end-of-course exams?
- Manual peer-assessment lays more burden on both students and teachers

The Way Forward

- Exploring the applicability of educational games
- Some positive results of introducing educational games in the physical sciences
- Although most studies focus on K-12 education
- Thorough reviews of educational games - Randel et al. (1992), Wu et al. (2012)
- CS advances may help with efficient integration of educational games into peer-assessment practices
- a way of eliciting participation through collaborative and competitive games

The Way Forward

All in all

- we still need robust design quality and measurement standards - still waiting for the first symposium on PA
- An opportune time for scholars in education and computer science to forge collaborations
- Not a practice within education anymore - **21st century PA is interdisciplinary**

References

All references can be retrieved from the article discussed in this talk

- Michael Mogessie Ashenafi (2015): Peer-assessment in higher education twenty-first century practices, challenges and the way forward, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2015.1100711