



Writing Assessment in the Laptop-mediated English Language Classroom: Rasch analysis, fairness and flexible delivery

Rob Miles 

Higher Colleges of Technology, United Arab Emirates
miles.rob@gmail.com

Abstract. This theory driven chapter documents a robust assessment strategy for writing in English that makes innovative use of technology and Rasch analysis to ensure highly efficient administration and fair and accurate grading of student work. The assessment strategy is flexible for both face to face and off campus administration, and offers an adaptable yet rigorous assessment tool that can survive the uncertainty of ever-changing delivery modes. The approach could be applied to any course seeking to assess student learning through written assessment such as essays or projects. This is an important consideration given the potential growth of distance and remote teaching worldwide in the wake of the Covid19 pandemic. The chapter contributes to on and offline formal written assessment, practical use of multi-faceted Rasch analysis, and highlights the impracticality of online proctoring software in this particular context.

Keywords: Writing Assessment, English Language, Rasch Analysis.

1 Introduction

Across the MENA region, the desire to move from ‘oil-reliant’ to ‘knowledge’ economies has seen governments and other influential actors invest heavily in education and technology [1]. In the UAE, the drive for educational excellence has seen English medium instruction (EMI) become the norm across K12 and tertiary education, where teaching occurs in a technology rich environment of 1:1 device deployment, Smart screens, online assessment and Learning Management Systems (LMS). Degree courses are largely taught in English, and students wishing to enter a degree program must meet an internationally benchmarked level of linguistic proficiency in order to obtain admission. For the federal institutions, this is currently a score of between 5.0 and 6.0 on the International English Language Testing System (IELTS), or the corresponding score on the nationally administered English proficiency test, the EmSAT. Students who do not achieve this can enroll on an English preparatory course in order to reach the required proficiency. This course is a prime example of the technologically rich teaching and learning environment typical to the UAE. Classroom devices – laptops – are

© The Author(s) 2023

A. Johnston et al. (eds.), *Proceedings of the HCT International General Education Conference (HCT-IGEC 2023)*, Atlantis Highlights in Social Sciences, Education and Humanities 13,
https://doi.org/10.2991/978-94-6463-286-6_6

deployed for each student, all materials are online and accessed through a Learning Management System (LMS), and assessments are carried out 100% online.

Success on the preparatory course is measured through a combination of course work assessments and final written examinations. The course does not employ an external examination (e.g. IELTS), nor the nationally recognized proficiency test (EmSAT). However, the validity and reliability of the preparatory course is recognized as robust by the institution, and consequently is accepted as proof of proficiency to progress to a degree program. A central pillar of the course's assessment strategy is testing students' written language skills through formal tests, worth 25% of the final grade. It is this key aspect that this chapter seeks to investigate.

Although writing assessments are delivered online through the LMS, the assessments take place in strict face to face environments. The COVID-19 emergency, however, necessitated the moving of all courses, and consequently the same assessments, into fully online delivery. While exerting enormous pressure and disruption on professional roles and practice [2] this appears to have taken place successfully. Writing exams were administered, students were assessed, and successful students were able to progress to degree programs in similar numbers to pre-COVID course delivery. This suggests that the writing assessments remained robust in terms of standards, but were flexible enough to adapt to the challenges of pandemic delivery modes. This deserves closer examination.

2 Research focus

Through the lens of Cultural Historical Activity Theory [3-5], three questions are addressed:

1. How is writing formally assessed in the laptop-mediated English language classroom?
2. How flexible is the assessment process in terms of on and off campus delivery?
3. What tensions exist within the assessment process?

The English preparatory course has created a robust yet flexible procedure for the formal assessment of students' writing skills in English. This procedure has been applied in both face-to-face traditional contexts, and online remote delivery, servicing 16 campuses and upwards of 1500 students. Technology is utilized at all stages from LMS delivery, to online marking and Rasch analysis and the delivery of a fair mark to students. Closer examination of this model may result in practices that can be shared and championed beyond the current context, and a research contribution to assessment in English for Academic purposes (EAP) and English Language Teaching (ELT) settings, formal written assessments, online assessment delivery and flexible delivery modes.

3 Literature Review

If the acquisition of a particular language skill is important, then it becomes necessary to assess the proficiency of that skill. Writing is no exception [6]. Indeed, it is only by assessing current writing ability that we can help students progress [7]. Key themes in writing assessment in this context are the impact of technology, manual and automatic evaluation of writing, assessment rubrics, Rasch analysis and online proctoring.

3.1 The impact of technology

Classroom technology advocates point to significant improvements in students' written abilities where devices are deployed in classrooms. In the USA, clear gains in standardized written assessments have been described in one-to-one laptop programs [8, 9]. These claims have been echoed in the UAE [10-12]. This is especially significant given the poor scores on written tests such as IELTS in the region. For example, the overall IELTS score in the UAE is 4.97, below the required entry for federal institutions [13]. If we are to measure these improvements then robust assessment procedures are necessary. The model under investigation utilizes technology at all stages, and if truly robust would allow valid and reliable measurement of claims as to the positive impact of technology on student writing itself, a 'positive tool for supporting student learning' [6].

3.2 Manual and automatic writing evaluation

Evaluating writing manually raises issues over quality and efficiency. Holistic single grading, for example as employed by the Test of English for International Communication (TOEIC), prevents raters from distinguishing between various aspects of writing including control of syntax, complexity of vocabulary, coherence or organization, and so on. There are also issues of consistency – one rater's 4 is not necessarily another's interpretation of a 4. Furthermore, analytic scoring based on "such features as content, organization, cohesion, register, vocabulary, grammar, or mechanics" is very time consuming [6]. The "brute force approach of 100% double marking" will achieve tight control over grades quality, but is "profligate, expensive and inefficient" [14]. An alternative to both is statistical marking, for example targeted reviews where experienced markers double check a sample of scripts [Raikes & Shaw, 2005 in 14].

Technological advances have seen a rise in automatic writing evaluation tools (AWE) that compare a student's work against a large database of writing in the same genre, and recent research has investigated the possibility of AWE feedback improving student writing [15-17]. Beyond this, The Pearson Test of English (PTE) employs AWE, and describes itself as a "breath of fresh air...without an examiner's bias" [18]. While research has shown good correlations between machine and human scoring, further pedagogical research is needed [19], and debate over the use and effectiveness of AWE is generally polarized between AWE as an effective tool or a refusal to believe that it can truly evaluate a piece of written work [20]. Indeed, AWE may not be a 'substitute for human intuition' [21]. The model in question may support the case for human evaluation.

3.3 Assessment rubrics

Writing is often evaluated through a rubric, or set of criteria, when manual marking is applied. Such rubrics may take the form of can-do statements, similar to those used in the Common European Framework of Reference (CEFR) [22]. For Shaw and Weir [14], the CEFR is best used as a ‘heuristic device’, rather than being prescriptive. The rubric used in the research site is an adaptation of the CEFR writing can-do statements. This rubric use mirrors internationally recognized writing assessments such as IELTS, where raters assign ‘bands’ to score a piece of writing. An alternative to rubrics might be comparative judgement. Comparative judgement presumes that an individual is more able to compare two performances than assign a score to a single one [23]. Basically, assessors compare pairs of scripts, deciding which one is better. This is repeated by different assessors until all the writing has been ‘compared’ several times by several assessors. The scripts can then be placed on a scale of performance that generates extremely reliable results of educational performance [24]. This may be an alternative to the ‘brute force’ of double marking using rubrics. However, ‘criterion-referenced’ marking using rubrics remains the norm [7].

3.4 Rasch analysis and writing evaluation

Rasch models, first proposed by George Rasch in 1960, are probabilistic measurement models using sophisticated mathematical procedures to calibrate parameters such as test-taker ability and item difficulty in an assessment setting [25]. Recent research examples include studies into the development of raters’ rating ability [26], and an investigation into rater severity at a university in Kuwait [27]. Rasch analysis has also been applied to AWE, suggesting that the automated system had greater consistency than human markers [28]. The application of Rasch models to writing assessment is of ongoing interest.

3.5 Online proctoring

Worldwide, online proctoring tools have generated significant controversy, as either the good proctor or big brother [29]. Questions also arise over ethical concerns with artificial intelligence in academic proctoring [30], with some arguing that ‘the story of online proctoring is difficult to disentangle from surveillance and policing’ [31], citing the need to create counter-narratives to the increasing adoption of such tools [32]. Online proctoring remains a polarizing subject.

4 Theoretical framework

This chapter takes activity theory as the theoretical framework. Rather than focusing on isolated elements activity theory allows consideration of the whole system.

Vygotsky [33] uses the following example. Imagine a student is trying to understand why water extinguishes fire. The individual elements, hydrogen and oxygen, are flammable and fire-sustaining. The extinguishing qualities of water are lost when you break the system into components [34]. Similarly, we cannot truly understand how an assessment is being designed and administered by considering isolated elements. Each element and complex interrelation needs to be considered in terms of the whole system [35]. The complex, multi-faceted process of assessment needs a complex theory that considers multiple elements. Activity theory has this utility. It provides the language and conceptual tools to describe and analyze the complexity of social situations like education [36-38]. There are 3 important concepts to consider.

4.1 The activity system

In activity theory, the unit of analysis is the activity system. An activity system could be a team, a department, an institution, a social system or practice. Regardless of size or scope, it is represented by the activity system [39]. See Figure 1.

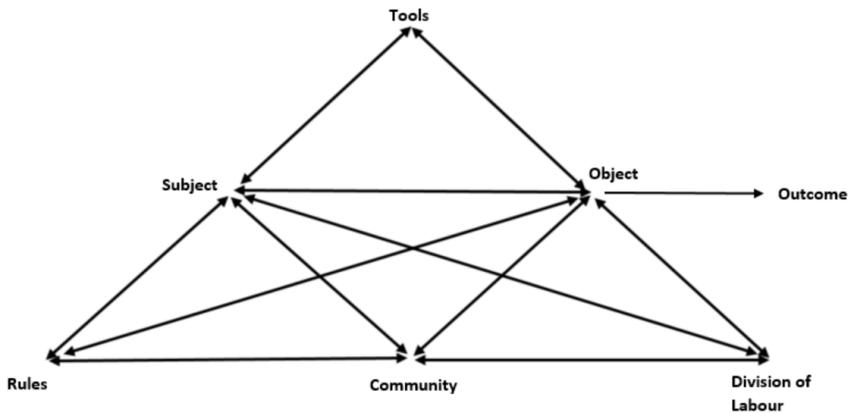


Fig. 1. the activity system as a node in a network of activity systems

All activity is social, and has an object with intended outcomes. The interaction between the subject (the individual, group or organisation) and object (the aim of the activity) is mediated by tools or instruments, which can be physical objects, concepts and ideas or social others. The activity system also includes rules, community and division of labour. Rules mediate between the subject and the community, while the division of labour mediates between community and object. In the example of a school or college, teachers (subjects) use instruments (course books, online quizzes, pedagogy) to teach their subject (object) with the intended outcome of students passing the course. There are rules of classroom behaviour, a wider community of students, teachers, parents and management, and expectations over who does what in and out of class (division of labour).

4.2 An inter-related network

Activity systems do not exist in isolation, but are in fact nodes in an inter-related network of systems. This is shown in Figure 2.

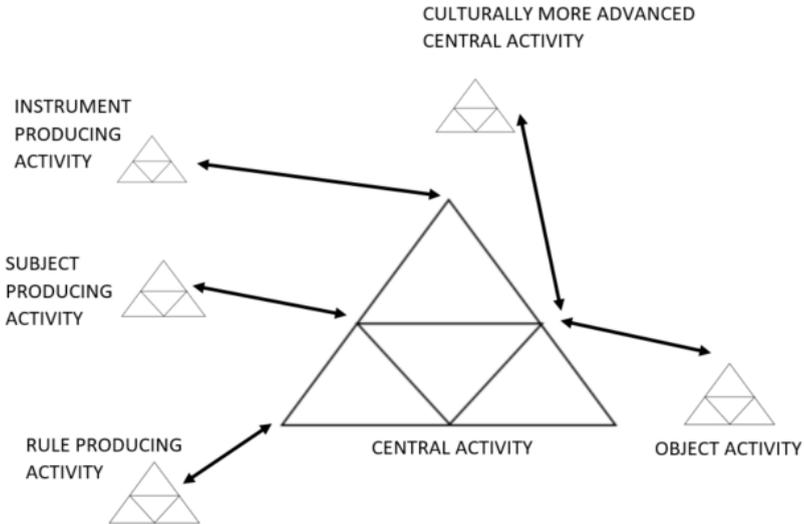


Fig. 2. the activity system as a node in a network of activity systems

Returning to the example of teaching, we have teacher training (subject producing activity), materials creation (instrument production), object activity (for example, assessment) and rule producing activity. There may also be a culturally more advanced version of the central activity itself. Developing nations, for example, might look to western education models as an ideal mirror for their own activity.

4.3 The principle of contradictions

All activity has an intended outcome, for example a school wants students to successfully pass their courses. The presence of unintended outcomes – for example, student failure and attrition – suggests the presence of contradictions. Contradictions are disruptions in, or between, nodes of the activity system, or between neighbouring systems in an inter-related network. The introduction of a new tool, for example a new course book, might lead to contradictions causing failure. A new technology might cause a negative reaction among teachers, again leading to failure in the intended outcome. Contradictions can be mapped to the activity system as shown in Figure 3.

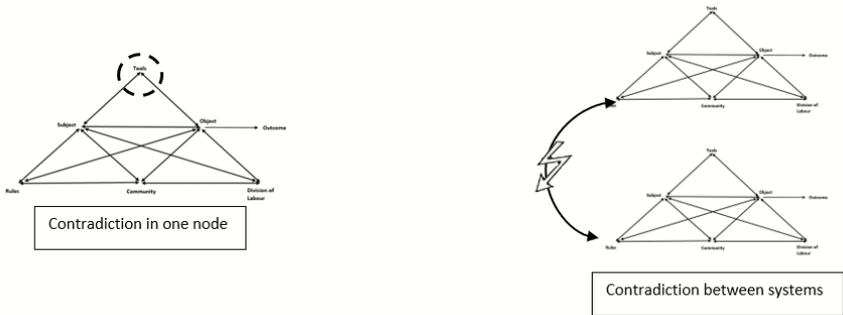


Fig. 3. Contradictions mapped to the activity system

It is important to identify contradictions. Solutions can then be applied, leading to the creation of new, more effective activity, through a process called expansive learning.

The framework of activity theory guides this project at all stages, from research design, to data collection, to analysis and presentation of results. The model of writing assessment is examined and presented in terms of an activity system and analyzed accordingly, as are tensions or contradictions and potential solutions.

5 Methodology

Activity theory is a qualitative framework, and qualitative research methods informed the data collection. The central activity of writing assessment is carried out by the Assessment Unit (AU). The AU members are the primary subjects for interview, although they remain anonymous in this report. Interviewees are referred to by letter and number (A1, A2 etc.) to avoid identification by gender or nationality. Once ethical approval had been obtained from the institution, all participants signed a consent form, and were able to withdraw participation at any time.

The interviews were based on Marken’s [40] 6-step model. See Figure 4.

What <i>tools</i> do the <i>subject</i> use to achieve their <i>objective</i> and how?
What <i>rules</i> affect the way the <i>subjects</i> achieve their <i>objective</i> and how?
How does the <i>division of labour</i> influence the way the <i>subjects</i> satisfy their <i>objective</i>?
How does the <i>tools</i> in use affect the way the <i>community</i> achieves the <i>objective</i>?
What <i>rules</i> affect the way the <i>community</i> satisfies their <i>objective</i> and how?
How does the <i>division of labour</i> affect the way the <i>community</i> achieves the <i>objective</i>?

Fig. 4. Marken's interview protocol

Semi-structured face to face or video conference interviews took place, and the recordings were then transcribed and analysed. Follow up interviews took place through email, and member checking of results was carried out.

6 Findings

The writing assessment process for the preparatory course is both straightforward yet complicated. It can be broken down into six clear inter-related activity systems, engaged to different degrees and complexity at the stages of pre-test, during-test and post-test. The six activity systems are

1. Assessment Creation Activity
2. Management Activity
3. Assessment Administration Activity
4. Test Taking Activity
5. Assessment Marking Activity
6. Assessment Grading Activity

6.1 Stage 1: pre-test

Assessment Creation Activity takes place prior to every assessment, and also involves Management and Administration. The five Assessment Unit (AU) members meet to choose suitable questions from a bank of prompts. This is a historical record of writing questions (or prompts) created by trained teachers following strict guidelines. One member of the AU explains

...we need to make sure that the prompts are things the students can write about [and] within their knowledge and experience...we don't want to make any of the prompts unfair...(A1)

Furthermore, the AU team strives to avoid

...bias towards either gender...or anything else...(A1)

The system is a 'streamlined process' that 'seems pretty robust' (A3).

Prompts are also checked for cultural appropriateness.

An Emirati, a member of the writing assessment committee, helps check the cultural/religious aspects and possible L1 influence that may cause confusion. (A4)

Teamwork is crucial as 'it shouldn't be an individual process' (A1). At this stage, the prompts are double-checked by management.

...it's good to have someone who's not a member of the AU, who hasn't gone through the process...[to] run their eyes over that...(A1)

Once approved, the prompts are sent to Administration. There are six prompts for any assessment that are distributed randomly to students via the Learning Management System (LMS). Administration configures the settings on the LMS including passwords, instructions, duration etc. as required. The tests are administered through a dedicated LMS course in which all current students are enrolled. Students with special Educational Needs (SEN) are enrolled separately with any required accommodations. Management then shares dates, times, instructions and passwords with individual campuses. For large administrations two (or more) sittings might be available. For each sitting, a unique bank of 6 prompts is used. An example prompt, as it appears in the LMS, can be seen in Figure 5.

Having a family is expensive nowadays. In your opinion, what are the causes of this, and what are some possible solutions?

You should write 250 words.

You have 40 minutes.

For the toolbar, press ALT+F10 (PC) or ALT+FN+F10 (Mac).

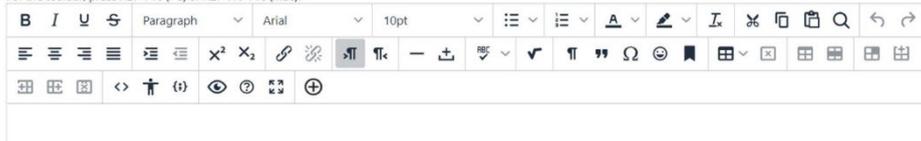


Fig. 5. Example writing prompt on the LMS

This stage is consistent for both face-to-face and 100% online assessment. For face-to-face administration, assessments are restricted to college Wi-Fi networks, and browser restricting software is employed. For 100% online assessments, there was no network restriction but online proctoring software was activated.

6.2 Stage 2: during-test

In Test Taking Activity, face-to-face writing assessment follows a traditional procedure. Students sit in exam rooms at separate desks. There are invigilators, or proctors, administering the exam and watching behaviour. Mobile phones, notes etc. are banned

from exam rooms. However, rather than writing on paper, students access the exams via the LMS. The exam is only accessible using a special browser that ‘locks’ students into one Internet window. The exams are password protected, restricted to college Wi-Fi networks and also limited by time and date. For example, an exam released for December 10th from 9:00 – 10:00 will only be available on that date between those times. Furthermore, the exams are timed and automatically submit after 40 minutes once started. Note that time can be extended for individual cases such as SEN students. Invigilators are expected to be active and vigilant during the exam, and to check that exams are submitted correctly when students finish.

100% online remote assessment during the pandemic followed a different procedure. Passwords were not used. Instead, students had to use an online proctoring system in order to access the assessment. The system used was built into the LMS and had a series of nine steps that students were required to complete before accessing the assessment:

1. Accept the terms of use
2. Check webcam and microphone
3. Read instructions
4. Read guidelines and tips
5. Take a photo
6. Show ID
7. Perform an environment check (point the camera around the room)
8. Facial detection
9. Begin the test

Students needed practice and training before sitting the assessments in order to be able to complete these steps. Failure to complete a step resulted in their being unable to access the assessment. As with face-to-face tests, students had 40 minutes to complete the assessment once started. A separate video conference room was made available for technical support.

6.3 Stage 3: post-test

Post-test, Administration extracts the student answers from the LMS, and uploads these to the Online Marking System (OMS) for Assessment Marking Activity. The OMS allows scripts to be assigned to markers, or raters, who are teachers currently teaching the course. Raters access the script and submit marks 100% online. Each script is double-marked by two raters. Management approves the release of scripts to raters and sets deadlines. Management also monitors the marking and keeps campus management apprised of progress. The OMS is a ‘very robust marking platform’ (A2) – raters do not receive their own students, have the option to leave comments and can also include links if they suspect that a script has been copied or memorised. The interface can be seen in Figure 6.

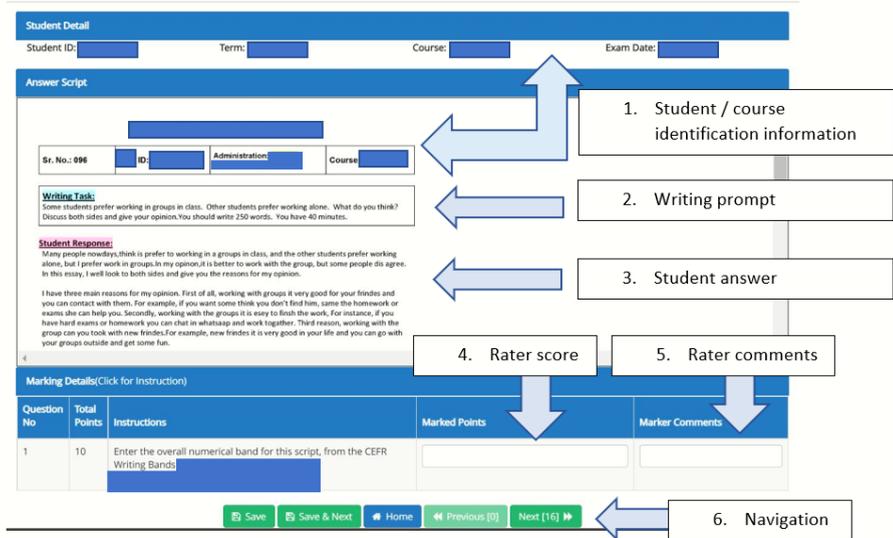


Fig. 6. The Online Marking System (OMS) interface

Raters use a rubric based on the CEFR can-do statements, called ‘bands’. The bands allow for analytical scoring, and have descriptors for General (overall), Vocabulary, Grammar, Cohesion and Mechanics. Raters give a holistic single band on the system. Figure 7 is an example of Band 7 (B1+) and Band 6 (B1) rubrics.

CEFR Writing Bands for English Communications & General English for Diploma

Band	CEFR	Written Production: Essays	Vocabulary Range & Control	Grammatical Accuracy	Coherece & Cohesion	Orthographic Control
7	B1+	<p>Can produce a short essay on a topic of interest, using simple language to:</p> <ul style="list-style-type: none"> list reasons, and give and justify an opinion. list causes/effects of a problem, and suggest solutions with some supporting details. list advantages/disadvantages with some supporting details. give a description with some supporting details. 	<p>Has a good range of vocabulary for familiar topics.</p> <p>Can produce appropriate collocations of some words.</p> <p>Has a sufficient range of vocabulary for topics of interest to explain the main points in an idea or problem with reasonable precision, but some repetition, incorrect word choice and errors still occur.</p>	<p>Communicates with reasonable accuracy in familiar contexts; generally good control, though with noticeable L1 influence.</p> <p>Uses some complex structures, but with limited accuracy.</p> <p>Errors occur, but it is clear what they are trying to express.</p>	<p>Can produce text that is reasonably organised and coherent, using a limited number of cohesive devices.</p> <p>Can introduce a counter-argument in a simple discursive text.</p>	<p>Can produce continuous writing which is mostly intelligible throughout.</p> <p>Layout and paragraphing are clear; spelling and punctuation are generally accurate.</p> <p>Almost no run-on sentences.</p>
6	B1	<p>Can produce a short, simple essay in a standard conventionalised format, which:</p> <ul style="list-style-type: none"> gives reasons and opinion on a familiar topic. gives causes/effects and solutions on a problem. lists advantages/disadvantages of a familiar topic. gives a description on a familiar topic. 	<p>Has a good range of vocabulary related to familiar topics and everyday situations such as family, interests, college, travel and current events, but lexical limitations may cause repetition.</p> <p>Shows good control of elementary vocabulary, but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.</p>	<p>Uses reasonably accurately a repertoire of frequently used patterns associated with more predictable situations or standard essay formats.</p> <p>Attempts complex structures, but with inaccuracy.</p> <p>Errors may cause some strain for the reader.</p>	<p>Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.</p> <p>Can form longer sentences and link them together using a limited number of cohesive devices.</p> <p>Can make simple, logical paragraph breaks in a longer text.</p>	<p>Can produce continuous writing which is generally intelligible throughout.</p> <p>Spelling, punctuation and layout are accurate enough to be followed most of the time.</p> <p>May have a few run-on sentences.</p>

Fig. 7. Example of Band 6 and 7 scoring rubrics

The rubric is widely accepted as robust and effective by those using it. The complete rubric has ten ‘bands’ corresponding to CEFR levels, and is able to grade students from C1 to below A1.

A zero grade is given when a response is “...completely off topic, or does not contain enough original language to accurately mark... [writing bands rubric]”. A memorised answer will likely be off topic. Similarly, language is well above the level may be unoriginal, and can often be found on google with a simple search. Such scripts are given zero, and a comment, for example “Script is off topic and seems memorised” or “Here is a link to the original essay online”.

Once all the marking is completed, Administration extracts all the data from the OMS for Assessment Grading Activity.

6.4 Rasch analysis

The use of multi-faceted Rasch analysis is a major differentiator of the preparatory program’s assessment process. This chapter does not set out to explain the principles behind Rasch analysis, but instead documents how the tool is used, how it is critical and how it effects student grades during Assessment Grading Activity.

Multi-faceted Rasch analysis is carried out using a program called Facets. Prior to data input into the program, the assessment unit first identifies any scripts that received zero. If both raters scored zero, ideally supported by comments such as a link to the original text, then the script is removed from the analysis. Raters should leave a comment with their zero grade as

... [it’s important] so we know why an essay has been given a zero...(A1)

True zero scores do not need to go through the process as there is no fair average score possible. A zero score means the text is unoriginal, or too short to rate, so it is not possible to generate a score. If only one rater has given zero, then the AU double checks the script and acts as a third rater. If the zero is merited, then the script receives a double zero score and is removed from the analysis. If the AU feels an actual score is merited, then this score replaces the zero and the script remains in the system for analysis.

Each rater, and each student, is given a unique identifier number. The input data for Facets essentially consists of student identifier, rater 1 identifier and score, rater 2 identifier and score. This is then put into Facets and the program carries out multi-faceted Rasch analysis. Facets uses sophisticated mathematical procedures, and is able to measure each rater’s leniency and consistency, and arrive at a fair average score for each student. A fair average score is a score that has been adjusted for rater severity or leniency, and is basically the score that student would have received if marked by two raters showing zero leniency or severity. Ideally, a rater will measure neutrally on leniency, being neither too strict or too lenient, and will also show consistency in the scores they are giving. Consistency is key as Facets can adjust for severity or leniency as long as the rater is consistent. Facets identifies ‘problem’ scripts as unidentified observations. This is caused when two raters who are normally severe give a lenient score, or lenient raters give a severe score. The AU will look at such instances and decide on an actual

fair score. As with zero scores, the assessment unit acts as a third marker and the final word on any issues. However, only a small number of scripts typically have issues identified by Facets. At the end of the process, each student has a fair grade that was neither disadvantaged by two overly severe raters, nor falsely advantaged by two extremely lenient raters.

Once the analysis is complete, scores are converted to percentages and returned to Management. Once Management confirms the grades, they are returned to Administration, and the final scores are then released to colleges and shared with the students. The assessment unit also prepares a report for the teachers. This includes a 'top raters table', populated by the best raters during that administration because

We feel that it's good that raters have an idea...I think people do get sort of a sense of pride or achievement... when they're on the table...(A1)

Indeed, one rater, on hearing they were number one let out a 'scream of delight' (A1). It seems the 'bragging rights' of being reported as a top rater are valued by the raters themselves. The process was essentially the same during 100% online remote assessment, although a far greater number of scripts received zero grades for suspected academic dishonesty.

Online proctoring software proved less than successful, and was not 'much of a deterrent' (A1). Some cases were obvious, for example a different person sitting the test, or clear communication with a third party. In other cases, it was more difficult to pinpoint any assessment malpractice. Many of the 'red flags' identified by the system proved little. Is a student looking at notes, or simply looking away from the screen? Is a student talking to themselves, or talking to someone in the room?

It [the online proctoring system] was a pain. It was very time consuming. And again, in the end, there's nothing we could do. Unless it was absolutely obvious, like someone else was doing it... (A1)

Even with the software, proving academic dishonesty remained very difficult especially given the nature of the course and assessment. For example,

With English, all you need is somebody who speaks English [to help you] ...but when it comes to... multivariable calculus, you need someone who knows how to deal with variable calculus (A3)

In other words, it is relatively easy to be helped with an English exam when compared to specialist subjects. Fortunately, the timing of the pandemic meant some classroom teaching had taken place, and therefore in-class samples of writing were available for comparison. Given the relatively large number of cases, this was then devolved to each campus. Writing committees were formed to look at each script in question, compare with the previous sample (where available) and then make a decision over scores. In many cases it was easy to identify the original source of copied writing. In other cases, the language of the essay (C1 or C2 on CEFR) was clearly well beyond the student's ability and previous sample. Where evidence was clear students received zero for the assessment. Where there was ambiguity, students could retake the test, but were warned about academic dishonesty.

7 Issues

For the assessment unit, the writing assessment is robust, reliable and well-honed. Valid, reliable tests are created, administered efficiently, marked in a timely fashion and students receive grades that reflect their writing ability. One member of the AU mentioned issues with time

Major exams like IELTS...go through like a piloting process...we can't do [that] (A3)

The exigencies of the course do not allow for piloting, but historical bank of prompts combined with analysis means that 'it's never a problem' (A3). Another member would like to see tighter links between the writing prompts and the course topics, and also mentions that typing skills need consideration

I have personally noticed that some students struggle more than others in terms of typing (A4)

but agrees that overall the process is 'very reliable' (A4).

However, issues do arise during Test Taking Activity and among raters during Assessment Marking Activity. Most prevalent during 100% online remote testing, some students also attempt to submit essays they did not write during face to face assessment. Through sophisticated procedures, such as third-party applications to bypass browser lock down systems, some students try to beat the system. The only way to combat this is through vigilant invigilation and equally vigilant marking leading to zero grades. For online assessment,

Some sort of online invigilation [is needed], where the teacher... can invigilate over zoom or something and keep an eye on the students (A1)

Such a system has yet to be tried out locally.

Rater issues relate to extreme severity and leniency, or lack of consistency. The Facets program identifies those who fall into these extremes, and reports go to Management. However, the process lacks a system to deal with consistently 'extreme' raters. The AU recognises the need to improve feedback, and give useful data 'in a way that's not going to make people upset' (A3). Extreme raters cause unexpected responses, and lead to more third marking, increasing workload and extending the grades turnaround time. One solution might be training, 'run a session...showing it's not all wizardry' (A1). Raters need to realise that being too severe, lenient or inconsistent, is not acceptable and is something that needs attention.

8 Results

The writing assessment process for the English preparatory course is straightforward and robust, yet makes complex use of Rasch analysis to ensure fair grades.

8.1 How is writing formally assessed in the laptop-mediated English language classroom?

The assessment process has three clear stages that involve six interrelated activity systems. These systems are involved to different degrees at each stage of pre, during and post-test. This interrelation and interaction are shown in Figure 9.

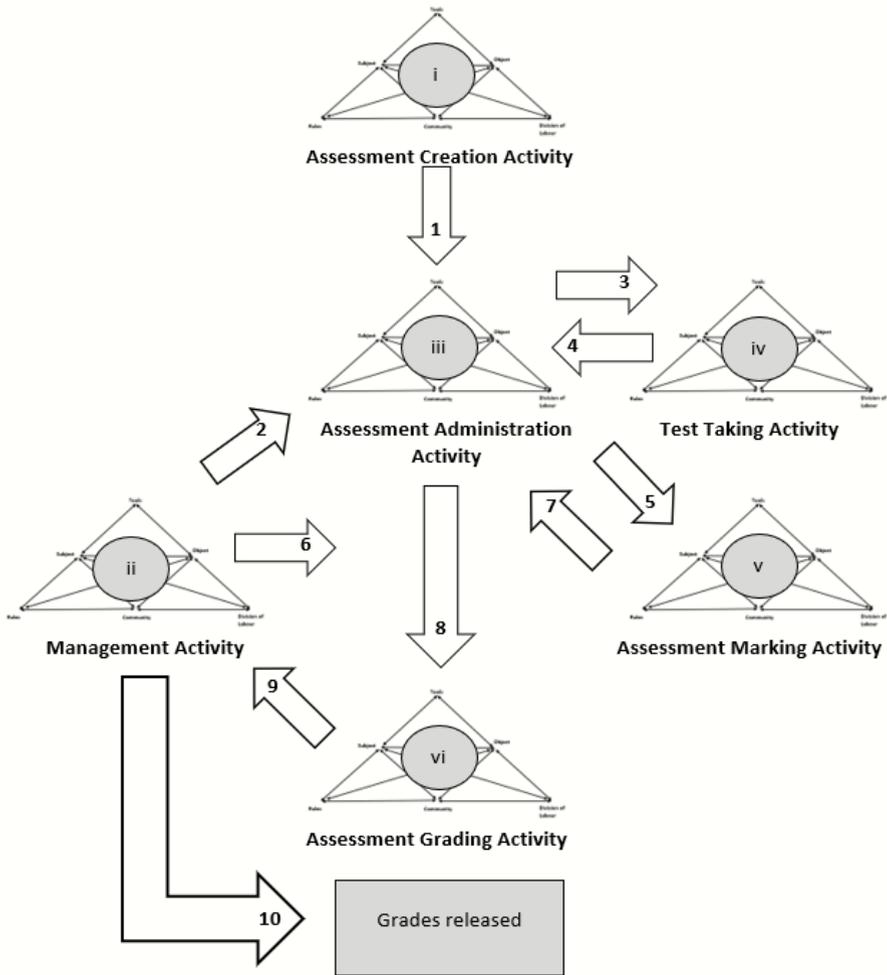


Fig. 8. Summary diagram of the interrelated activity systems during assessment

- KEY**
- 1a. Assessment Unit chooses 6 prompts.
 - 1b. Management sets assessment date and confirms prompts.
 2. Administration creates assessment on LMS. Assessment is released and carried out on given date / time.
 3. Post-test, Administration extracts scripts from LMS.
 - 4a. Administration uploads scripts to Online Marking System (OMS).
 - 4b. Management sets marking deadline and monitors marking activity.
 5. Marked scripts return to Administration from the OMS.
 6. Administration sends marked scripts to the AU for grading.
 7. AU sends final grades to Management for approval.
 8. Management approves final grades, and grades are released.

Writing Assessment as one activity can be mapped as shown in Figure 9. The Assessment Unit use tools, namely the LMS, the Online Marking System, Facets (for Rasch analysis) and their own assessment expertise with the object of valid, reliable assessments giving students fair grades. There are clear rules assessment administration, and a community of students, college management and teachers. The labour is then divided between the six interrelated systems.

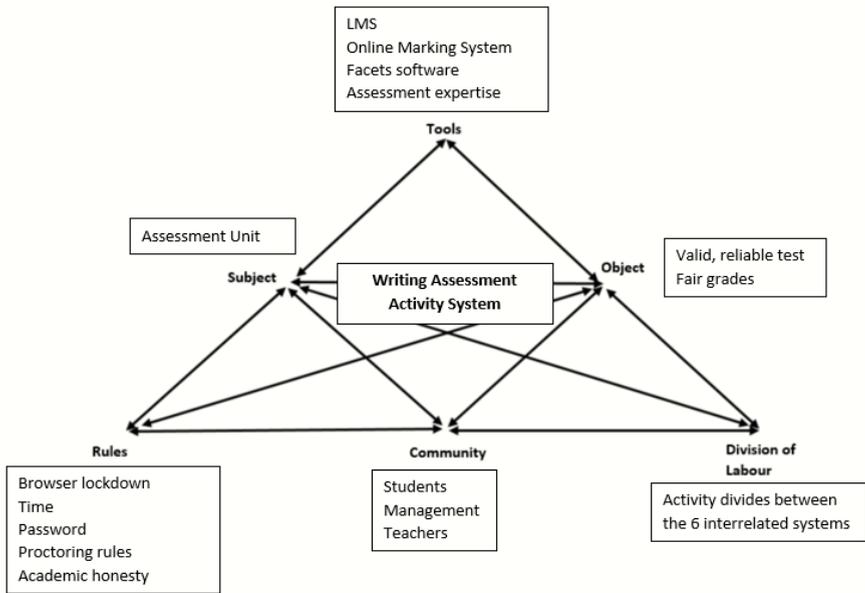


Fig. 9. The Writing Assessment Activity System

8.2 How flexible is the assessment process in terms of on and off campus delivery?

The system is extremely flexible. Assessments can be carried out in traditional face to face settings, or delivered online through the LMS without any major changes to the process. Diligence is needed during marking to identify scripts that are not the students’ own work, however. Online proctoring software has not been effective.

8.3 What tensions exist within the assessment process?

There are two major tensions in the process. Firstly, some markers are overly severe or lenient, and action is needed to address this issue. This is mapped to the activity system as Contradiction 1 in tools. Secondly, some students try to submit work that is not their own, especially during 100% online administration. This is mapped to the activity system in Figure 10 as Contradiction 2 in rules.

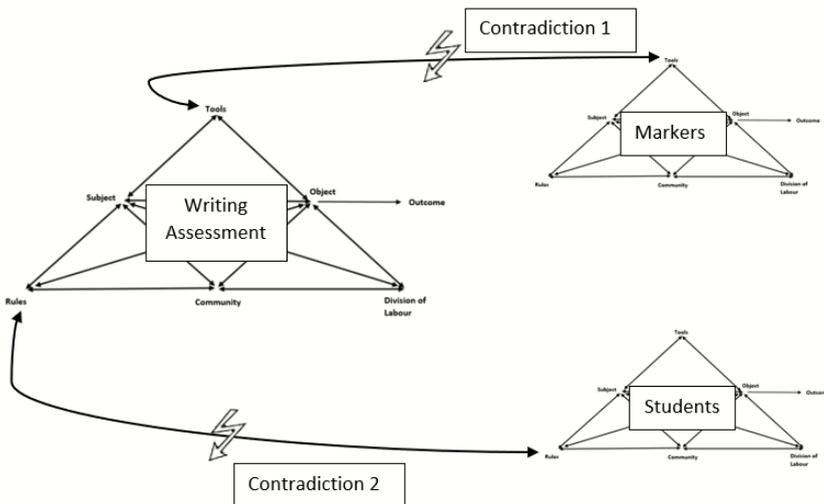


Fig. 10. Contradiction 1 and 2

9 Discussion

Technology is a key component in the writing assessment process for the English preparatory course. The assessments are delivered online through an LMS, whether administered face to face or remotely. Scripts are submitted online, and marking also takes place online. Facets software is utilized to carry out Rasch analysis and ensure fair student grades. Writing is a key skill, and therefore it is necessary to assess it [6].

Technology enhances the whole process, ensuring valid assessments, smooth administration and fair grades. While the Assessment Unit do not mention any advantages in terms of the teaching of writing, teachers on the course know that their students are assessed fairly. Teachers are made aware of their leniency (or otherwise) and consistency, and it would be interesting to investigate further what, if any, washback this has into classroom teaching. This could support those claiming improvements to writing as a result of classroom technology [8-12]. A robust assessment procedure is able to accurately measure writing ability, and thus becomes a 'positive tool for supporting student learning' [6].

The process does employ the "brute force" approach of double marking [14], but has the advantage over external examinations such as IELTS in that it does not need to 'pay' its markers. The teachers – markers - are already employed by the institution, and marking takes place during working hours, thus removing cost issues. The use of Rasch analysis allows for the monitoring of marking standards and the fair 'balancing' of student grades. Without Rasch, a student could be unfairly advantaged or disadvantaged. A student with a true writing band of 5 might have two severe markers who score their work as 4. On the other hand, two lenient markers might score the script a 6. By using Facets software harshness and leniency can be identified and moderated. Similarly, situations with disagreement, for example marker 1 gives a 6 and marker 2 gives a 4, can also be identified, and moderated. By giving the markers feedback on their performance there should also be improvement in the standard of marking leading to less severity, leniency and inconsistency. Not only is the procedure robust, but the standard of marking is also monitored to ensure high standards. Statistical marking [14], to replace double marking through targeted review, is a future possibility, but is unfortunately 'limited...by our system (OMS)' (A3). The AU does, however, recognize that this is a path worth pursuing.

Another strength of the process is the marking rubric. Although grades are holistic, there are diagnostics for grammar, vocabulary, coherence and orthographic control, employing the CEFR as a 'heuristic device' [14] that means students writing ability is firmly anchored to a measured and recognized scale with 'explicit and systematic scoring criteria' [7]. One member of the AU mentioned the possibility of using AWE, but recognizes that 'the software is not there for us yet' and 'it could get quite complicated and expensive' (A3). AWE is not yet a viable option, and human evaluation remains key [21]. Similarly, comparative judgement [23, 24], as an alternative to rubrics and double marking, is something 'we disagree on' (A3) and thus remains at an early pilot stage.

The procedure is flexible, and was able to efficiently switch to 100% remote testing during the pandemic. However, online proctoring software was not successful. The interviewees did not demonstrate dystopian views of online proctoring as big brother [29], ethical concerns [30] or make comparisons to police states [31], but instead take a more practical counter-narrative. Except in obvious cases the software did not work, and the task of checking 'red flags' was onerous, time consuming and ineffective. Academic dishonesty was dealt with instead through vigilant marking procedures. For 100% remote assessment, alternatives to online proctoring software should be found.

This experience of the impracticality of the software in this context is a valuable contribution to the counter-narratives [32].

The strength of the whole system is shown by the fact that only two contradictions are identified. Inconsistent marking is identifiable, with further training suggested as a solution. The problem of academic dishonesty is perhaps thornier, but the interviewees were confident in the ability of the system and markers to identify most cases, and procedures are in place to deal any issues.

10 Conclusion

The Writing Assessment activity and process described here is robust, uses technology effectively, and guarantees accurate and fair assessment for all students. It could provide a model for other institutions with the same technical capabilities and available resources. The model is a good example of how Facets and Rasch analysis can be used in a practical context, and further research could examine the effect and washback this has on the actual teaching of writing. The model also shows the impracticality of online proctoring software in this context, and suggests that one solution might be live online proctoring, perhaps using video conference software. Finally, the complex activity described here demonstrates how the overall activity of writing assessment is in fact a series of interrelated activity systems that contribute to different stages of the process. The activity already suggests solutions to the twin contradictions of marker unreliability and student academic dishonesty. Solving either issue would only strengthen the process further. Writing is an important skill for students in English medium academic environments, and therefore it is important to assess that skill. It is equally important that this assessment is robust, valid, and produces fair grades for students that reflect their true ability. These are the main strengths of the writing assessment process described here.

11 References

1. Miles R, Al-Ali S, Charles T, Hill C, Bligh B. Technology enhanced learning in the MENA region: Introduction to the Special Issue. *Studies in Technology Enhanced Learning*. 2021;1(2):293-300.
2. Watermeyer R, Crick T, Knight C, Goodall J. COVID-19 and digital disruption in UK universities: Afflictions and affordances of emergency online migration. *Higher Education*. 2021;81(3):623-41.
3. Engeström Y. *Learning by expanding: An activity-theoretical approach to developmental research*. 2nd ed. New York, NY: Cambridge University Press; 1987/2015.

4. Engeström Y. Developmental studies of work as a testbench of activity theory: The case of primary care medical practice. *Understanding practice: Perspectives on activity and context.* 1993:64-103.
5. Engeström Y. Activity theory and the social construction of knowledge: A story of four umpires. *Organization.* 2000;7(2):301-10.
6. Weigle SC. *Assessing writing:* Cambridge University Press; 2002.
7. Hyland K. *Second language writing:* Cambridge university press; 2019.
8. Grimes D, Warschauer M. Learning with laptops: A multi-method case study. *Journal of Educational Computing Research.* 2008;38(3):305-32.
9. Park Y, Warschauer M. Syntactic enhancement and second language literacy: An experimental study. *Language Learning & Technology.* 2016;20(3):180-99.
10. Mokhtar IA, Al Bustami G, Elnimeiri A, editors. *Use of Weblogs by Non-native EL students: A case study in the UAE. E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education;* 2009: Association for the Advancement of Computing in Education (AACE).
11. Raddawi R, Bilikozen N. ELT Professors' Perspectives on the Use of E-rubrics in an Academic Writing Class in a University in the UAE. *Assessing EFL Writing in the 21st Century Arab World:* Springer; 2018. p. 221-60.
12. Tubaishat A, Bataineh E, editors. *The impact of laptops on female students learning in an outcome-based Learning environment. Society for Information Technology & Teacher Education International Conference;* 2009: Association for the Advancement of Computing in Education (AACE).
13. Ielts. *Power BI Report.* 2020.
14. Shaw S, Weir C. *Examining Writing: Research and Practice in Assessing Second Language Writing (Studies in Language Testing 26).* 2007.
15. Yamashita T. *Corrective feedback in computer-mediated collaborative writing and revision contributions.* 2021.
16. Lv X, Ren W, Xie Y. The effects of online feedback on ESL/EFL writing: a meta-analysis. *The Asia-Pacific Education Researcher.* 2021;30(6):643-53.
17. Riezebos Z-TZ, Ming-Hua Y, Peter. *A research framework of smart education. Smart Learning Environments.* 2016;3(1):1-17.
18. Pearson. *Pearson English Language Tests 2022* [Available from: <https://www.pearsonpte.com/>].
19. Ware P, Kessler G. *CALL and digital feedback. Contemporary computer-assisted language learning.* 2013:323-41.
20. Hockly N. *Automated writing evaluation.* Oxford University Press UK; 2019.
21. McCallum L, Coombe C. *Examining the Cycle of Assessing Writing in MENA Contexts: Trends, Challenges and Best Practice Recommendations. European Journal of Applied Linguistics and TEFL.* 2021;10(1):79-99.
22. Division CoECfCC-oECML. *Common European framework of reference for languages: Learning, teaching, assessment:* Cambridge University Press; 2001.
23. Lesterhuis M, Verhavert S, Coertjens L, Donche V, De Maeyer S. *Comparative judgement as a promising alternative to score competences. Innovative practices for higher education assessment and measurement:* IGI Global; 2017. p. 119-38.

24. Pollitt A. Comparative judgement for assessment. *International Journal of Technology and Design Education*. 2012;22(2):157-70.
25. Barkaoui K. Multifaceted Rasch analysis for test evaluation. *The companion to language assessment*. 2013;3:1301-22.
26. Yan X, Chuang P-L. How do raters learn to rate? Many-facet Rasch modeling of rater performance over the course of a rater certification program. *Language Testing*. 2019;02655322221074913.
27. Erguvan ID, Dunya BA. Analyzing rater severity in a freshman composition course using many facet Rasch measurement. *Language Testing in Asia*. 2020;10(1):1-20.
28. Chan KKY, Bond T, Yan Z. Application of an Automated Essay Scoring engine to English writing assessment using Many-Facet Rasch Measurement. *Language Testing*. 2022;02655322221076025.
29. Coghlan S, Miller T, Paterson J. Good proctor or “big brother”? Ethics of online exam supervision technologies. *Philosophy & Technology*. 2021;34(4):1581-606.
30. Henry JV, Oliver M. Who Will Watch the Watchmen? The Ethico-political Arrangements of Algorithmic Proctoring for Academic Integrity. *Postdigital Science and Education*. 2021:1-24.
31. Logan C. Toward abolishing online proctoring: Counter-narratives, deep change, and pedagogies of educational dignity. *Journal of Interactive Technology and Pedagogy*. 2021;20.
32. Selwyn N, O’Neill C, Smith G, Andrejevic M, Gu X. A necessary evil? The rise of online exam proctoring in Australian universities. *Media International Australia*. 2021:1329878X211005862.
33. Vygotsky LS. *Thought and language-Revised edition*. Cambridge, MA: Massachusetts Institute of Technology. 1986.
34. Virkkunen J, Newnham DS. *The Change Laboratory: A tool for collaborative development of work and education*: Springer Science & Business Media; 2013.
35. Miles R. Making a Case for Cultural Historical Activity Theory: Examples of CHAT in Practice. *Studies in Technology Enhanced Learning [Internet]*. 2020; 1(1).
36. Bligh B, Flood M. Activity theory in empirical higher education research: choices, uses and values. *Tertiary Education and Management*. 2017;23(2):125-52.
37. Hopwood N, Stocks C. Teaching development for doctoral students: what can we learn from activity theory? *International Journal for Academic Development*. 2008;13(3):187-98.
38. Murphy E, Rodriguez-Manzanares MA. *Activity theory perspectives on technology in higher education*: IGI Global; 2013.
39. Engeström Y. *Learning by expanding: An activity-theoretical approach to developmental research*. 2nd ed. New York, NY: Cambridge University Press; 1987.
40. Marken JA. An application of activity theory. *Performance Improvement Quarterly*. 2006;19(2):27-49.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

