Z E S Z Y T Y N A U K O W E UNIWERSYTETU RZESZOWSKIEGO

SERIA FILOLOGICZNA
ZESZYT 117/2022 STUDIA ANGLICA RESOVIENSIA 19

DOI: 10.15584/sar.2022.19.4

Przemysław KRAKOWIAN
University of Lodz, Poland

przemyslaw.krakowian@filologia.uni.lodz.pl

INVESTIGATING USER GROUP BEHAVIOR PATTERNS IN DIGITAL EXAMINATIONS

Abstract: Computer Mediated Communication (CMC), as well as Electronically Mediated Communication (EMC), is becoming a staple feature of many working environments. More recently, Computer Mediated Assessment (CMA) in the form of Testing Assessment Systems (TMS) has joined the ranks of arrangements where learning, information exchange and assessment take place in digital landscapes, where on top of deploying assessment, they constitute a meeting ground for teachers, assessors, and students/learners. Any such system has the potential to offer numerous other opportunities including amongst others a training platform, a system for creating learner portfolios, a vetting system for rating scale construction or a multimedia annotated corpus of learner language. Additionally, information stored in a TMS may be used for research purposes, collaborative arrangements for professional development, vetting and training solutions. This paper presents what has emerged from developing procedures and functionalities in digital exams, and what may appear as attractive opportunities offering insights into the functioning of digital assessment and digital testing environments. While, admittedly, functionalities presented here have been gradually, and while in most cases, added ad hoc to the system, over the course of eleven years since the environment was first put into operation, have come, not only to complement one another, but also to function as a coherent system. In their current state, they offer practical solutions in quality control, training, research, and systems forensics.

Keywords: language education, language testing, digital assessment, technology, pandemic

Introduction

What is now known as Computer Assisted Testing Systems at the Institute of English Studies at the University of Lodz, or CATS (cf. cats.uni.lodz.pl), also lovingly referred to as *Koty* (Polish for feline) by staff and students, was first developed in the academic year 2011/2012 as a computerized placement test to process incoming cohorts of students and place them at the levels appropriate to

their ability to replace the paper-based placement test and to streamline the procedure. Originally there was very little intent and interest to deploy it large scale as replacement for the in-house, paper-and-pen assessment in the English language programme. The original computerized placement test was unproctored and delivered as a distributed arrangement, which meant that it could be accessed at leisure and convenience via the Internet, providing the placement information before the start of the study programme, a feature particularly useful for incoming foreign students. The primary consideration in developing the system was a logistical one, and it operated on a very straightforward principle of deploying the test content, storing the answers, and providing a placement result through automated scoring. The test comprised mostly closed items such as multiple choice, true-false or multiple matching items, with very few open questions, where the choice of answer was severely limited through context or instructions, so that the scoring key would carry only a small and finite number of answers. Some modifications to the key were introduced to account for issues that were spotted over the first few years of operation. One very puzzling alteration to the scoring procedure involved what must have resulted from transfer of training – capitalization that persistently appeared in the answers that was perhaps drummed into the examinees by their former teachers, as that may have been a prominent legibility issue in paper-and-pen tests. The test in its original form, with some minor modifications reflecting the need to be accessed on a variety of mobile devices has been in continuous operation for 11 years, and in fact the first background functionality that was introduced in the system was that to register and store information on the type of device, screen resolution, type of browser and later also IP-geopositional data.

Digitized Assessment vs. Testing Management System – C.A.T.S.

In the year 2015, after relocating to a new campus, and following the pilots designed to test student reactions as well as to work out technical issues in the early 2015, the CATS were deployed full scale in the winter semester exams of the academic year 2015/2016. This was possible owing to the saturation of the new campus with, amongst other, high-capacity computer facilities allowing to process large cohorts of students. Until now, the system has been used to process 39 exam events for the total of approximately 9,000 students, involving the joint efforts of over 90 staff in the roles of task and item developers, examiners, markers, raters, invigilators, monitors and test marshals. Over the years a simple computerized test has progressed to a fully-fledged Testing Management System (TMS).

The early add-ons to what was essentially a digitized exam used to deploy content and communicate results, involved anti-fraudulent measures, ranging from a variety of access codes and two step authentications, with burner, one-time codes,

to white-listing IP addresses for access to different parts of the system. The latest additions involve linking the system to anti-plagiarism solutions, and tracking ways in which the answers are provided and altered, particularly in the essay-type writing tasks where students type their essays online, and markers mark the essays using a facility called Script Assist, also online, often from the comfort of their own homes.

Figure 1. Script Assist – online marking tool for evaluating written performance

Some newer, pandemic-driven functionalities follow the workflow implemented for marking writing, to assist rating speaking via the use of what is referred to as EPSS or Electronic Performance System, deployed on mobile devices now handed out for the exam on tablets, although with large screen phones this is also a viable solution. In the pandemic and remote examinations, the EPSS was configured to use Sharepoint and MS Teams recording facility to allow the raters to input the results directly to the system. The obvious training potential of such materials, both for assessing writing and speaking papers is later discussed in gaining insight to examiner behavior patterns.

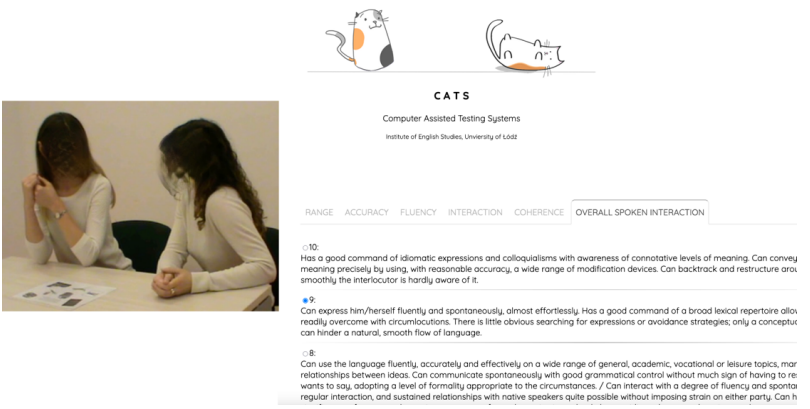

Figure 2. EPSS deployed on mobile devices to allow real-time input of grades to TMS – (portions of image blurred in accordance with GDPR)

Online and digitally mediated learning and assessment environments in the structure of Learning Management Systems, Virtual Learning Environments and ePortfolia, as well as Testing Management Systems, all of which may be considered to be some implementation of Communities of Practice (Wenger, 1999; Wenger, McDermott, & Snyder, 2002; Gadomska & Krakowian, 2017), where such a community (CoP) is often defined as a network or an arrangement through which ideas are exchanged and solutions generated, are all fairly well established in the mainstream educational practice. It often is a community of like-minded people associated with one another through similarity of interests and goals, and working on a common set of problems, in common pursuit of solutions, and themselves creating a store of knowledge (Wenger, White, & Smith, 2009). Such community of practice additionally entails the process of social learning that takes place when individuals who have common interests in some field or problem collaborate and share ideas, come up with solutions and otherwise interact with each other to work (Saint-Onge and Wallace 2003, Hildreth and Kimble 2004; Blackmore, 2010).

In such environments, largely depending on their character and purpose, considerable amounts of activity data are stored, some never to be accessed, and there merely for the record, some instrumental in grading, some for communicating progress or for the purpose of assessment. Some, paradoxically, have been prompted to a large extent by the need to monitor behavior and aimed at establishing content and delivery appropriateness, as well as any examinee behavior that might be questionable or fraudulent in nature. The information available through ICT procedures implemented in computerized tests that can be used to safeguard the fairness of the measures as well as to prevent examination material from being compromised by unauthorized access, unsanctioned use of resources, or unendorsed technology, comes in the form of user behavior logs. Depending on how detailed they are conceived to be, they may either merely contain log-in and log-off data, IP geo-positioning, or come in such abundance that it is possible to recreate every keystroke and every click of the mouse.

While useful in gathering evidence of foul play, they may equally be instrumental in establishing other types of behavior; one such functionality was deployed to track examinee preferences regarding how the text was deployed in reading comprehension tasks. Traditional pen-and-paper exams use a full-page display of texts, which sometimes spans over more than one page. Such arrangement is somewhat difficult to implement in computerized tests, as there is always a compromise between fitting the text on a single screen, and text size and ensuing legibility. In order to overcome this issue, a carousel mode of text display was originally used, with individual paragraphs, or larger chunks of text, that were rotated at a click of the mouse. After receiving some negative feedback from the examinees, a toggle between carousel and full page display was introduced.

While they may not realize this, the examinees, with every click of the mouse, with every singular stroke of the keyboard, leave a digital fingerprint in the data capture systems behind digitally mediated learning environments. Under normal circumstances, this data is used to administer the systems and store the end product of the learner in the form of assignments, essays, blogs and numerous other activities, from the more traditional to the more socially and collectively oriented. With some modifications, the data capture systems, together with visualization procedures, will allow to present a complete picture of all the attempts, modifications and revisions before the end product is achieved. In the case of the toggle carousel view to full page view, information gathered to establish which of the modes was in fact used by students, as implementing various functionalities is time and resource intensive, indicated that as novelty wore off, the majority of students in fact opted for the carousel mode, and it is now predominantly used. In the case of investigating how the students arrived at the end product through editing, interesting data has been collected that may shed light into exam taking and writing strategies, yet an unexpected offshoot was that during the time when the exams were taken remotely from homes, under somewhat lax supervision offered by MS Teams and web cameras, some dishonest behavior was identified, as the process of producing written work was recorded by the system alongside the final product.

Activity tracking in the COVID-19 pandemic

The Polish regulations concerning legitimacy of computerized assessment outlined in the Journal of Law (2007 No. 188, section 1374, with subsequent amendments) make provisions as to when computerized assessment as a way of ascertaining students' competencies and knowledge is permissible by law, and if so what conditions must be met so that they are not to be considered low-stake assessments and instrumental as formative rather than summative assessments (Drasgow & Olson-Buchanan 1999, Drasgow 2015). During normal operation in pre-pandemic conditions, exam monitors were a solution to meeting such mandatory regulations, with an additional arrangement for special circumstances where safe exam browser plugin was used (Heintz 2017) for events when students would arrive with their own computers. Such events were conceived to provide additional exam opportunities beyond monitored computer labs, and SEB would make cheating impossible, as, in essence, it shuts access on the user's device to the hard drive and any applications, leaving only the computerized test active. Unfortunately, SEB also disables the camera and microphone that under the pandemic have extensively been used to monitor examinee behavior during the test. To overcome this limitation alternative functionalities were implemented.

Behind the frontend CATS interface that the examinees see, a red flag is raised every time the system encounters a certain type of behavior defined as suspicious, such as attempts to access a segment of the system from illegal locations, or by a person who does not have sufficient privileges to interact with that segment, both by examinee as well as staff. All of those flags are stored alongside the user id and with a time stamp and can be compared with other activity in the same timeframe or segment of the system to determine if this was merely an offhand chance incident, or whether it requires reporting to appropriate regulating authorities. Other flags include text being pasted rather than typed, and artefacts that in such instances are often copied alongside the text unbeknown to the student, mostly formatting information if the text is copied from a word processor or a website. This was a provision that additionally steamrolled what is referred to as the *digital divide*, which is defined as the gap between certain segments of society, who do not have access to, or whose access to computers and computer-like devices and/or the Internet is limited (Pachler, Bachmair, Cook & Kress 2009, Parsons 2012). The largest contributing factor here is of economic nature, but as Pachler et al. (2009) point out, other reasons such as age, place of habitation, level of education and social standing, as well as gender have been identified. Unfortunately, the conditions brought upon by the pandemic seem to have extended this notion of the digital divide to include what has been recognized as the either real connectivity issues or simulated underhand examinee behavior to avoid monitoring. This new dimension of the digital divide appears to be the most significant obstacle to monitoring digital assessments live via video streaming, if they are to be considered proctored exams as per the aforementioned legal validity considerations.

Investigating marker behavior patterns in online assessment of writing

Biber, Johansson, Leech, Conrad and Finegan (1999) show is how coordination and subordination are used to express meaning in texts. Indices involving the total number of coordinate and subordinate phrases in student writing can be easily computed, with the assumption that more advanced learner language will be characterized by a larger subordination index (SI) and lower coordination index (CI). Alongside keywords, sentence length, CI and SI ratios can be regarded as stable parameters instrumental in identifying authorship of texts, and while there is no automatic procedure for deploying those measures in the CATS, they may be used to deal with any red-flagged behavior detected by the system. Type to Token ratios (TTR) are one of the contributing parameters, where tokens in a text are defined as individual words, and in WST they can be equated with the number of running words in the text (McEnery and Wilson 2001), while types, on the other hand,

refer to classes of words, where a class comprises any variant or lemma of the word or the word itself that is repeated in the text. Keywords, on the other hand, refer to lexical items whose frequency in the text are elevated when compared to the reference body of text, usually large, to allow statistical inferencing. Keywords tend to re-occur in texts written by the same person as a result of lexical predilection which they reflect in terms of lexical idiosyncrasy, as opposed to a large collection of normalized texts in a reference corpus (McEnery and Wilson 2001).

Each essay entry in CATS is accompanied by author id, that can be referenced to a particular text, and series can be exported for analysis. Other than SI, CI, TTR, the notions of propositional density and idea density are an attractive premise in investigating examinee idiolect and may be used alongside WST indices. The CPIDR (Computerized Propositional Idea Density Rater, pronounced "spider") is a computer program that allows the researcher to establish the propositional idea density of a written or transcribed spoken text without human intervention (Brown, Snodgrass, Kemper, Herman and Covington 2008). The authors of the program claim that it has been validated against human raters and the convergence is sufficiently high to lead to further applications in machine aided assessment (MAA).

Propositional density, also known as proposition density, or P-density, but sometimes referred to as propositional idea density, and understood as in Kintsch (1974) and Turner and Greene (1977), can be determined by the total number of content words such as verbs, adjectives, adverbs, prepositions, and conjunctions against by the total number of words (Snowdon, Kemper, Mortimer, Greiner, Wekstein and Markesbery 1996). The measures obtained through the CPIDR constitute a gauge of the effectiveness and economy of expression when referring to concepts and ideas, with the notion that while complex ideas require complex language, complexity can also be attributed to economy of means. This measure is not implemented as an automatic feature of CATS, but is a semi-automatic feature drawn upon, when, post hoc features are used to validate the exam performance.

In short, if a red flag is thrown on an essay, where measures obtained automatically diverge from marker evaluations, third, and sometimes additional parties are called in to resolve the issue. While this is far from an arrangement where the essay obtains a fair mark from the start, such iteration allows for human errors, be it inter, intra-rater reliability issues, or numerous intervening issues.

Obviously, there is great potential to what directions this inspection and intervention takes route.

Investigating rater behavior patterns in online assessment of speaking

While some studies that address the assessment of speaking in exam contexts suggest that raters may not feel as comfortable assessing pronunciation as they

do other aspects of a speaker's performance (Orr 2002, Hubbard, Gilbert and Pidcock 2006, Brown 2006, De Velle 2008), more recent investigations of rater behaviour involving electronic evidence from training, maintenance and online examination programmes tentatively show that pronunciation, in fact, is the first category examiners attend to (Hubbard 2011, Chambers and Ingham 2011, Krakowian 2011, Seed 2012, Tynan 2015).

Most evaluation schemas involve provisions for handling assessment of pronunciation ranging from intelligibility and accurate production of individual sounds, through managing word and sentence stress and appropriate intonation, to such use of phonological features that they convey and enhance meaning. It is interesting, however; to look at what happens when examiners need to make ratings of oral expression in the absence of explicit scales to handle assessment of pronunciation.

The observations recorded in CATS make use of a large batch of pre-tested and in some cases standardised samples of oral expression with different assessment schemas and raters that were assigned duties in the exam following their previous training, involvement or teaching duties to make a claim that what sounds *nice*, *good*, and *pleasant* (all adjectives used by examiners in their comments) may sometimes obtain more merit than it actually deserves, and that female raters rate more leniently, moreover; female raters rate male speakers more favourably and tend to disregard errors, and, finally there is a tendency to evaluate more favourably those whom the raters are familiar with. The data for this claim comes from hard evidence registered in CATS EPSS from before the pandemic and, to a lesser, degree, during the pandemic.

No	*Measure*	*MN.SQ*	*t-fit stat.*	*Overall S/L*	*Leniency*	*Severity*	*No*
1	1.375633	0.7142857	-0.1951748	-2.197225	0	**2.197225**	***1***
2	0.7611303	1.785714	0.4415075	-1.386294	0	1.098612	***2***
3	-1.375633	3.928571	**1.600518**	-2.079442	0	**2.079442**	***3***
4	-0.7611303	3.214286	1.261665	1.94591	**1.94591**	0	***4***
5	-0.2450292	2.5	0.8771017	-2.484907	**1.609438**	**1.94591**	***5***
6	1.375633	3.571429	**1.623343**	-2.197225	**1.386294**	**1.609438**	***6***
7	1.37563	3.571443	**1.643233**	-1.791759	0	**1.609438**	***7***
8	1.375633	1.428571	0.2130144	-1.098612	0	1.098612	***8***

Figure 3. Analysing behaviour patterns in assessment of recorded samples – anchoring samples to identify leniency and severity in assessment.

The data pertaining to the portion of samples accumulated in the CATS was subjected to analysis using the Multi Facet Rasch Analysis to confirm the postulated and identified trends and, if possible, to identify additional trends if any. Before performing MFRA Chi-square statistic used on the premise that behaviour departing from the postulated model, whatever model that may be, is penalised

by the statistic in the form of a residual – represented as t-fit. Depending on the direction, it is considered as underscoring or overscoring, indicative of too severe or too lenient assessment respectively. The data is logged into a html file for easy inspection online, with problem areas indicated in bold. The calculations are performed by the system, as they are relatively straightforward. In the case of the logit based statistic in Rasch Analysis, the residual is essentially based on the same principle, though the procedure is infinitely more complex and involves the application of the exponential function (for discussion on Rasch Analysis consult Wright and Stone 1979, Wright and Masters 1982, Wilson 2005, Bond and Fox 2007, Krakowian 2010). The research assumption beyond it was that there would be some overlap permitting to confirm already identified trends and to identify additional processes or additional samples that conformed to the patterns identified earlier.

Obsvd Score	Obsvd Count	Obsvd Average	Fair-Z Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	N. Rater
61	17	3.6	3.41	1.36	.39	.09	7	**2.0**	5	.00	1
55	17	3.2	3.35	.79	.49	0.3	-2	0.3	-2	.98	2
5	17	3.3	3.43	-1.36	.49	**4.5**	-1	**8.4**	-1	.98	**3**
25	16	1.6	1.21	0.71	.70	1.3	4	.6	4	.11	4
56	17	3.3	3.43	-.25	.49	**1.7**	-2	**2.5**	-1	.98	**5**
63	18	3.5	3.51	1.34	.46	**1.9**	8	**1.6**	6	.15	**6**
58	17	3.4	3.61	1.35	.49	**1.7**	-3	1.2	-2	.99	**7**
93	17	3.5	5.89	1.35	.66	.3	4	1.0	9	.11	8
Obsvd Score	Obsvd Count	Obsvd Average	Fair-Z Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	PtBis	N. Rater

Figure 4. Behaviour patterns processed in FACETS – data identified by the system in Figure 1

The Multi Facet Rasch Analysis procedure, performed using FACETS (Wright and Stone 1979, Wright and Masters 1982) a Many-Facet Rasch Analysis dichotomous and polytomous model program, did indeed confirm the existence of the groups postulated earlier in the a priori analysis. Since the program is executable code and not open source, it has to be run manually and post hoc, and in most cases is used as a redundant measure after the completion of assessment.

Concluding remarks

Even though originally conceived as measures to aid exam proctoring, as it required by the laws and regulations, the various functionalities and inspection perspectives may have the potential to unveil what is happening between the screen, keyboard and the system. This process naturally will demand new tools to be developed, amongst others for similar purposes as it was first done with CATS. And while this is an expected and inevitable direction, there is no good

reason to assume that the various developments of the system will in a direct fashion offer new insights into how exams are taken in the digital ecology of learning and education, how different user groups can be investigated; for whatever reasons might be viable for the stakeholders and user groups.

References

Biber, Douglas, Johansson, Stig, Leech, Goeffrey, Conrad, Susan, and Finegan, Edward (1999). *Longman Grammar of Spoken and Written English.* Pearson Education Limited.

Blackmore, Chris (2010). *Social Learning Systems and Communities of Practice.* Springer London.

Bond, Trevor, and Fox, Christine., 2007, *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. University of Toledo Press

Botley, Simon, Mcenery, Tony and Wilson, Andrew (2000). *Multilingual Corpora in Teaching and Research*. Editions Rodopi

Brown Cati, Snodgrass Tony, Kemper Susan, Herman Ruth and Covington, Michael (2008) Automatic Measurement of Propositional Idea Density from Part-of-speech Tagging, *Behavioral Research Methods*; 40(2): 540–545.

Brown, Annie., 2006, An examination of the rating process in the revised IELTS Speaking Test., *IELTS Research Reports* Volume 6, IELTS Australia, Canberra and British Council, London, pp 41–65

Cohen, Andrew (2014). *Strategies in Learning and Using a Second Language.* Taylor & Francis

De Velle, Sacha., 2008, The revised IELTS Pronunciation scale., *ESOL Research Notes*, vol 34, pp 36–38

Drasgow, Fritz. 2015. *Technology and Testing: Improving Educational and Psychological Measurement.* NCME Applications of Educational Measurement and Assessment Book Series. Taylor & Francis.

Dror, Itiel (2011). *Technology Enhanced Learning and Cognition.* John Benjamins Publishing Company.

Gadomska, Agnieszka, & Krakowian, Przemysław. (2017). Implementing Blogs for Developing Academic Writing Skills in a Variety of Higher Educational Contexts. *International Journal of Continuing Engineering Education and Life-long Learning*, 27(4), 1.

Gardenfors, Peter, & Johansson, Petter. (2014). *Cognition, Education, and Communication Technology.* Taylor &Francis

Gilquin, Gaëtanelle & Paquot, Magali, (2008). Too Chatty: Learner Academic Writing and Register Variation. *English Text Construction* 1(1):41–61.

Gledhill Christopher, (2000), Collocations in Science Writing. *Language in Performance Series* no. 22, Tübingen, Gunter Narr Verlag.

Heintz, Alexander (2017). *Cheating at Digital Exams – Vulnerabilities and Countermeasures*. Norwegian University of Science and Technology Department of Computer Science

Hildreth, Paul, & Kimble, Chris (2004). *Knowledge Networks: Innovation Through Communities of Practice.* London Hershey: Idea Group Inc.

Hoc, Jean-michel, Cacciabue, Pietro, & Hollnagel, Eric (2013*). Expertise and Technology: Cognition & Human-computer Cooperation.* Taylor & Francis.

Hubbard, Christopher., Gilbert, Susan., and Pidcock, John., 2006, Assessment processes in speaking tests: a pilot verbal protocol study., *ESOL Research Notes*, vol 24, pp 14–19

Journal of Law 2007 no. 188, Section 1374: Rozporządzenie Ministra Nauki i Szkolnictwa Wyższego z Dnia 25 Września 2007 r. w Sprawie Warunków Jakie Muszą Być Spełnione, aby Zajęcia Dydaktyczne na Studiach Mogły Być Prowadzone z Wykorzystaniem Metod i Technik Kształcenia na Odległość, dz. u. 2007 nr. 188, poz. 1374 z Późn. Zm.

Kintsch Walter (1974) *The Representation of Meaning in Memory*. Erlbaum.
Koh, Hiam Seng, & lee, sai Choo. (2008). *Information Communication Technology in Education: Singapore's ICT Masterplans, 1997–2008*. World Scientific Publishing Company PTE Limited.
Krakowian, Przemysław. 2010. *Modern Test Theory Explained*. Warszawa: Oficyna Wydawnicza Scholar.
Krakowian, Przemysław. 2018. *Classical and Modern Test Theory in Developing and Validating Language Tests*. Wydawnictwo Naukowe PWSZ w Płocku
Leila, Barbara, and Scott, Mike eds. (1994) *Reflections on Language Learning.* Multilingual Matters
Li, li (2017). *New Technologies and Language Learning*. Macmillan Education UK.
Mcenery, Tony and Wilson, Andrew (2001). *Corpus Linguistics: An Introduction*. Edinburgh University Press
O'Neill, Brien, & Gillespie, Alex. (2014). *Assistive Technology for Cognition: A Handbook for Clinicians and Developers*. Taylor & Francis.
Orr, Michael., 2002, The FCE Speaking test: using rater reports to help interpret test scores., *System*, vol 30, no 2, pp 143–154
Pachler, Norbert, Ben Bachmair, John Cook, and Gunther Kress. 2009. *Mobile Learning: Structures, Agency, Practices*. Springer US.
Parsons, David. 2012. *Refining Current Practices in Mobile and Blended Learning: new Applications: new Applications. Advances in Mobile and Distance Learnin.*g IGI Global.
Saint-Onge, Hubert, & Wallace, Debra (2012). *Leveraging Communities of Practice for Strategic Advantage.* Taylor & Francis.
Seed, George., 2012, Perceptions of authenticity in academic test tasks., *ESOL Research Notes*, vol 49, pp 17–26
Seok, Shoonhwa, Meyen, Edward & Boaventura Dacosta (2010). *Handbook of Research on Human Cognition and Assistive Technology: Design, Accessibility and Transdisciplinary Perspectives: Design, Accessibility and Transdisciplinary Perspectives.* Medical Information Science Reference.
Snowdon David, Kemper Susan, Mortimer James, Greiner Lydia, Wekstein David, and Markesbery William (1996). *Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life: Findings from the nun Study.* JAMA 528–532.
Turner Althea & Greene Edith (1977). *The Construction and use of a Propositional Text Base (Technical Report 63)*. University of Colorado, Institute for the Study of Intellectual Behavior, Boulder.
Tynan, Robert., 2015, Creating ePortfolios to facilitate and evidence progress using learning technologies., *Cambridge Exams Research Notes*, vol 61, pp 55–68
Wenger, Etienne (1999). *Communities of Practice: Learning, Meaning, and Identity.* Cambridge University Press.
Wenger, Etienne, Mcdermott, Richard, & Snyder, William (2002). *Cultivating Communities of Practice: AGuide to Managing Knowledge.* Harvard Business School Press.
Wenger, Etienne, White, Nancy, and Smith, John (2009). *Digital Habitats: Stewarding Technology for Communities.* CPSquare
Wilson, Mark., 2005, *Constructing Measures: An Item Response Model* Lawrence Erlbaum Associates
Wright, Benjamin., and Masters, Goeffrey., 1982, *Rating Scale Analysis*. MESA Press
Wright, Benjsamin., and Stone, Mark., 1979, *Best Test Design*. MESA Press