

CALIBRATION IN EFL READING: EXAMINING THE NATURE OF JUDGMENT OF CONFIDENCE

*Madhu Neupane**

ABSTRACT

This article examines the nature of English as a foreign language (EFL) learners' appraisal confidence and appraisal calibration in EFL reading comprehension. Appraisal confidence refers to the degree to which test takers identify the probability that their test answer is correct or appropriate in percentage terms (e.g., 0%, 25%, 50%, 75%, or 100%). Appraisal calibration refers to the accuracy of test takers' appraisal confidence by comparing their appraisal confidence to their test performance in percentage terms. Two hundred and three students studying Master of Education (M. Ed) with specialization in English in the Tribhuvan University participated in the study. An EFL reading comprehension test specially designed for the study and appraisal confidence rating scales incorporated in the same reading comprehension test were used as the tools for data collection. The findings of the study show that the students were highly overconfident in their reading comprehension because the difference between their average appraisal confidence (86.84) and average accuracy in performance (52.35) was +34.49. The implications of the study and recommendations for further research are discussed.

Key Words: EFL reading comprehension, appraisal confidence, appraisal calibration

In the present day knowledge-based economy (Kumar & Welsum, 2013) reading is one of the main sources of knowledge. Grabe (2009) highlights the need for modern citizens to be skilled readers when he writes, “[S]uccess is much harder to come by without being a skilled reader” (p. 5). English is a global language. Therefore, EFL reading plays a significant role in advanced studies, academic success, cross-cultural awareness, economic and professional competition as well as active and meaningful participation in modern societies (Grabe, 2009; Grabe & Stoller, 2011). However, an alarmingly large number of adult EFL learners are unable to pursue their goals because of their inability to comprehend complicated texts (Berne, 2004). For this reason, the question of how learners' reading abilities can be improved is of great importance.

* Ms. Neupane is Lecturer in English Education at Central Department of Education, T.U., Kirtipur, Nepal

Successful reading comprehension is usually defined as the reader's understanding of the message expressed by the writer (Nuttall, 2005). Understanding a message is a complex process which involves processing texts at lower levels (e.g., at the lexical, syntactic, and semantic levels) as well as at higher levels (e.g., understanding the overall organisation of the text, interpreting the text according to the situation and context, using background knowledge, and making inferences). Though the execution of the lower level processes can be automatised with extensive practice, learners' metacognitive aspects play a significant role at the higher level processing of information (Block, 1992; Casanave, 1988; Grabe & Stoller, 2011; Mills, Pajares, & Herron, 2007). Among different aspects of metacognition, comprehension monitoring plays a crucial role in reading comprehension as it helps the readers to evaluate and regulate their own ongoing comprehension process (Baker, 1979; Han, 2012). This comprehension monitoring aspect of metacognition is also said to be reflected in learners' judgements of their appraisal confidence (usually measured by using appraisal confidence rating scales ranging from 0% to 100%) in likelihood of the accuracy of their performance (Kleitman & Stankov, 2007, Phakiti, 2016). The overall scores of such judgements can also be used to make comparisons between learners' overall appraisal confidence scores and their accuracy in their task performance that is calibration of their performance.

APPRAISAL CONFIDENCE AND ITS MEASUREMENT

Appraisal confidence refers to the degree to which test takers identify the probability that their test answer is correct or appropriate in percentage terms (e.g., 0%, 25%, 50%, 75%, or 100%). Test takers assess their appraisal confidence using available information about the perceived difficulty of the test task and how well they think they have performed in answering a given question.

Appraisal confidence-rating scales are embedded in each test question, so that test takers can report on their appraisal confidence immediately after they have answered a test question (Phakiti, 2016). Test takers can be asked to rate the extent to which they believe in the correctness of their responses to test items and tasks in percentage terms. For example, they may be asked to indicate their belief that a response is correct with a 0%, 50%, or 100% probability (e.g., Björkman, 1994; Yates, Lee, & Shinotsuka, 1996). Alternatively, researchers may ask students to quantify their appraisal confidence as high, medium, or low (Glenberg & Epstein, 1987). In the present study, appraisal confidence is estimated by using ratio scales of appraisal confidence in percentage terms as shown in the following example:

Task 2: Tick (✓) the best alternative. **Rate your appraisal confidence as you answer each question.**

1. Which of the following does **NOT** describe a scientist?
 - (a) They give consideration to the possible replication of their work.
 - (b) They control other people in the working environment.
 - (c) They enjoy their work.
 - (d) They do not want to talk to people who do not respect their work.

0%	25%	50%	75%	90%	100%
----	-----	-----	-----	-----	------

APPRAISAL CALIBRATION AND ITS MEASUREMENT

Appraisal calibration refers to the accuracy of test takers' appraisal confidence by comparing it to their test performance in percentage terms. If appraisal confidence and test performance match, test takers are said to be well calibrated. It is then hypothesized that a test takers' ability to accurately appraise their performance will ultimately lead to better performance and the likelihood of future learning success (Phakiti, 2016). Calibration expresses the correspondence between subjective and objective probability, that is a relative frequency (Bjorkman, 1992).

According to Phakiti (2016), calibration can simply be computed by subtracting the rated percentage of appraisal confidence from the percentage of actual test performance at the item level or at an overall level. On the basis of this calculation method, test takers are said to be well calibrated when their appraisal confidence level matches their test performance (Harvey, 1997; Jonsson & Allwood, 2003; Kleitman & Moscrop, 2010; Phakiti, 2016). This occurs when the appraisal calibration score is zero. For example, if on average a test taker reports a 50% appraisal confidence and the actual test performance is also 50%, the test taker is considered well calibrated ($50 - 50 = 0$). In contrast, the mismatch between the accuracy of judgement and objective accuracy is referred to as miscalibration (Maclellan 2014; Phakiti, 2016; Stankov, Pallier, *et al.*, 2012). Similarly, the score of under or over confidence rating is called bias score (Morony, Kleitman, Lee, & Stankov, 2013; Pallier *et al.*, 2002) or realism score (Stankov & Lee 2008). For example, if the appraisal confidence is 75%, but the test performance is 50%, the test taker is said to be overconfident. If the appraisal confidence is 25%, but the test performance is 50%, the test taker is said to be underconfident. Research has shown that people are typically overconfident in the judgement they express (Arkes, Christensen, Lai, & Blumer, 1987, p. 133). A group

calibration diagram can be used to present test takers' appraisal calibration vividly so that readers can understand it readily.

Figure 1 presents an example of an appraisal calibration diagram. The 45° line (called a unity line) represents the test performance. If the appraisal confidence rating is on the 45° line, the examinee is calibrated. If the appraisal confidence rating is above the 45° line, the examinee is overconfident; if below the 45° line, the examinee is underconfident. Ideally, an appraisal calibration score should be on or close to the unity line. Taking errors of measurement and the probabilistic nature of appraisal confidence into account, an examinee has good appraisal calibration when the appraisal calibration value is within $\pm 5\%$, rather than the exact 0% (Phakiti, 2016).

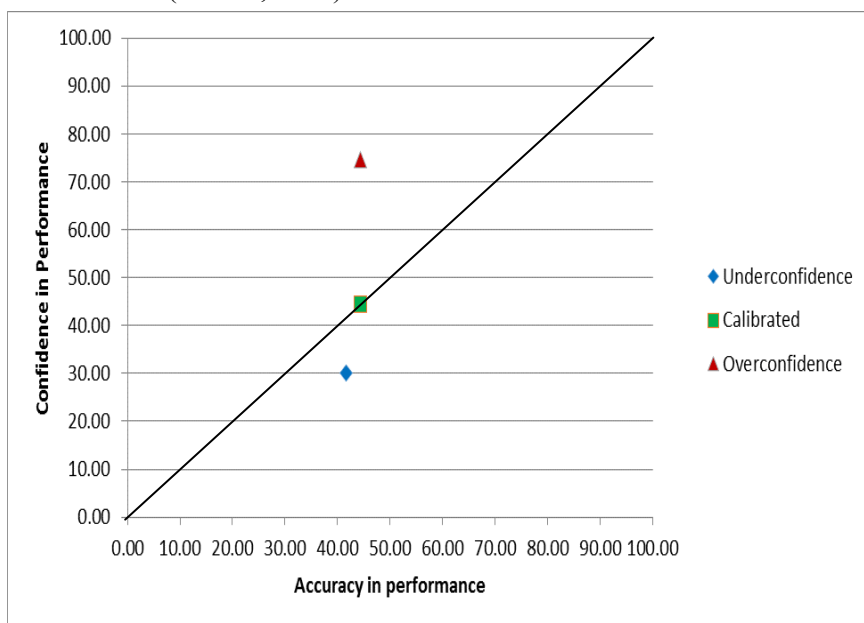


Figure 1: Appraisal calibration diagram for an individual

REVIEW OF RELATED LITERATURE

Research on calibration has a rich history in the past (Dinsmore & Parkinson, 2013). Previous research on calibration has focused on methods of improving calibration (Arkes *et al.*, 1987; Hacker, Dunlosky, & Graesser, 2009). Arkes *et al.*'s (1987) research involving two experiments regarding reducing the overconfidence of undergraduate students showed that by providing feedback on learners' performance (answer to individual questions they answer) and asking for justification to their answer, learners' calibration can be improved. Hacker *et al.*'s (2009) quasi-experimental design involving 137 college students

investigated the impact of extrinsic incentives and reflection on students' calibration of exam performance; and the relationships among attributional style, performance, and calibration judgments. The findings of the study showed that higher-performing students were very accurate in their calibration. However, lower performing students were less accurate in their calibration, and students in the incentives condition showed significant increases in calibration. The qualitative data also revealed differences by performance level in open-ended explanations for calibration judgments.

Further research has focused on nature of learners confidence (Dinsmore & Parkinson 2013; Hadwin & Webster 2013). Dinsmore and Parkinson's (2013) study on 72 (11 males and 62 females) university level students' calibration in reading using Bandura's (1986) model of reciprocal determinism showed that the participants level of calibration was at acceptable level and participants based their confidence ratings on prior knowledge, characteristics of the text, characteristics of the item, guessing and combinations of these categories. Similarly, Hadwin and Webster (2013) examination of the nature of confidence judgments associated with personal goal setting of 170 students enrolled in a first-year undergraduate course indicated that judgments of confidences were better calibrated with self-evaluations of current goal attainment than past goal attainment, learners did become less overconfident during the experiment of nine weeks, and learners who were performing better at university tended to be better calibrated. Pervious research has also investigated the role of prior knowledge (van Loon, de Bruin, van Gog, & van Merriënboer 2013) to primary-school children's (n = 103) commission of errors and overconfidence in these errors when learning new concepts. Findings indicated that inaccurate prior knowledge affects children's learning and calibration as children were found more overconfident and less receptive to concepts from further study when they had activated inaccurate prior knowledge.

A very recent study by Phakiti (2016) explored the nature and relationships among test takers' performance appraisals, appraisal calibration, and reported cognitive and metacognitive strategy use in a language test situation. Two hundred and ninety-four English as a foreign language (EFL) students took an English test, which was designed to measure four language areas (listening, grammar, vocabulary, and reading). The students reported their level of appraisal confidence immediately after answering each test question. At the end of the test, they were asked to report their overall appraisal confidence and perceived cognitive and metacognitive strategy use in the test. The findings indicated that test takers were not well calibrated in all test sections; their

appraisal confidence could predict just above one third of the test performance variance; they tended to be underconfident in easy questions but overconfident in difficult questions; and appraisal calibration was not strongly related to reported metacognitive strategy use.

The review of previous research shows that research on calibration of EFL reading comprehension performance has gain sufficient attention. The present research was conducted to address this knowledge gap.

RESEARCH QUESTIONS

The present study aims to address the following research question:

- What is the nature of EFL learners' appraisal confidence and appraisal calibration for the EFL reading comprehension performance?

METHODOLOGY

SETTING AND PARTICIPANTS

This research was carried out at the Department of English Education of a Nepalese university located in Kathmandu. The participants of the study were the students studying for the Master of Education (M.Ed.) degree with specialisation in English. About 210 participants took part in the study but usable data came from just 203 participants due to incomplete data. Out of 203 participants, 115 (56.65%) were males and 88 (43.35%) were females. They were between the ages of 20 and 32 (mean = 24.06), and had studied English for between 10 and 25 years at the time the research was carried out (mean=15.23).

RESEARCH INSTRUMENTS

In order to answer the research question, an EFL reading comprehension test and appraisal confidence scales incorporated in the reading comprehension test were used. For the purpose of this research, a reading comprehension test was designed as per the specification of skills and strategies that constitute reading. Two expository texts were selected from Van Doren (1992) and Richardson (2010). Expository texts are informational texts which are usually written in the present tense and use a high number of technical words. They may be organised sequentially, make comparisons, underline contrasts, or describe cause and effect. They may also be descriptions and exhibit complex structures, often within the same paragraph or passage (Akhondi, Malayeri, & Samad, 2011). Such texts were chosen because university students are required to read a significant number of such texts as a part of their academic program. Similarly, the texts had a difficulty level of grade 13+, the basic level for university students, according to Fry's (1977) readability formula.

Reading comprehension is said to comprise a number of skills and strategies (Alderson, 2000; Grabe & Stoller, 2011; Nuttall, 2005). For this research, the specifications of reading skills adapted from Phakiti (2009) were used. The skills are described as follows:

- *Identifying factual information*: Readers are required to locate and identify answers to questions about specific information and details in the passage. In such questions, both the question information and correct answers are found in the text.
- *Making inferences*: Readers are required to draw conclusions based on the information in the passage. Such questions require general knowledge, skillful reading, and higher order processing of information on the part of the readers.
- *Getting the meaning of vocabulary in context*: Readers are required to identify the meaning of a word or phrase as used in the passage.
- *Identifying the main idea (of the text), purpose, attitude or opinion (of the writer)*: Readers are required to identify the main idea or the subject of the whole passage, or the author's attitude or opinion towards the content or main purpose of the text.
- *Identifying references*: Readers are required to identify antecedent (a word, phrase, or a sentence) to which a pronoun or other expression refers.
- *Retrieving elliptical information*: Readers are required to retrieve the information that is deleted from the text by using the context.

In order to test all these skills, a reading comprehension test must adopt questions of different kinds, thereby requiring readers to use a variety of the skills listed above. Moreover, a suitable reading comprehension test must satisfy other criteria, including reliability and suitable length given the time available for participants to complete the test. The test used in this research was therefore piloted with 15 EFL learners. At the piloting stage, the test contained 50 questions, each worth one point. Based on the information obtained from the pilot regarding the time taken for the participants to complete the test, the clarity of expression in the questions, and the contribution of each item to the overall reliability of the test, fourteen items were deleted, and other items were modified. The finalised reading comprehension test consisted of 36 items each worth one point and the reliability (Cronbach alpha) of the test was 0.75. The test techniques used in the test are described as follows:

- *Multiple-choice questions*: In this type of question, readers are provided with a stem and four alternatives from which they have to choose the correct answer.

- *True/false/not given*: In this type of question, readers are provided with statements and are asked to identify whether the statements are true or false according to the information provided in the text. If the text does not contain enough information to answer ‘true’ or ‘false’ they are asked to choose ‘not given’.
- *Matching*: In this type of question, readers are asked to match the given words with their (contextual) meanings.
- *Filling in the blanks*: In this type of question, readers are asked to supply missing information that can be recovered from the context.
- *Identifying referents*: In this type of question, readers are asked to identify the words, phrases or sentences that the underlined reference words (such as he, she, it, which, etc.) refer to in the given text.
- *Short answer questions*: In this type of questions, readers are required to provide a brief answer—one or two sentences. Multiple answers are accepted if they demonstrate an understanding of the text. In this research, for this type of questions, the answer key was prepared based on the pilot of the EFL reading comprehension test. Table 1 summarises the test techniques.

Table 1: Summary of test techniques in the EFL reading comprehension test

Reading texts	Types and number of questions						Total
	True, False, Not Given	Multiple choice	Filling in the blanks	Identifying references	Matching	Short answer question	
Text A	4	4	-	3	3	3	17
Text B	3	5	2	3	3	3	19
Total	7	9	2	6	6	6	36

APPRAISAL CONFIDENCE RATING SCALES

Theoretically an appraisal confidence rating scale depends on the number of alternatives (k) given to a multiple choice question (i.e., $100/k$) (Phakiti, 2016). However, in the EFL reading test designed in the current research, some of the questions had alternatives while others were open ended. Therefore, the same six-point relative frequency appraisal confidence scale (i.e., 0%, 25%, 50%, 75%, 90%, and 100%) (Phakiti, 2016) was used for all the questions irrespective of question types. The relative frequency appraisal confidence scale was embedded into each test question. The questions were designed to allow the learners to record both their answers and appraisal confidence estimates. The participants were

instructed to rate their appraisal confidence immediately after they answered each question.

DATA PREPARATION AND ANALYSIS

Three main steps were taken in preparing the data collected from the reading comprehension test and appraisal confidence rating scales. First the data were entered into SPSS version 22. The scores of reading comprehension test and appraisal confidence rating scales were used for data analysis. In SPSS spreadsheet, a test score and its appraisal confidence were paired in the data entry. The descriptive statistics of each data set were first computed to check whether the normal distribution assumptions were met. The reliability and internal consistency of all the research instruments were calculated by using Cronbach’s alpha coefficient. In order to address the research question raised in the study (the nature of EFL learners’ appraisal confidence and appraisal calibration for the EFL reading comprehension test) first the students’ scores in the EFL reading comprehension test were converted into percentages and descriptive statistics (minimum, maximum, mean, standard deviation,) were examined.

RESULTS AND DISCUSSION

RESULTS OF PRELIMINARY ANALYSIS OF THE RESEARCH INSTRUMENTS

Table 2 presents the descriptive statistics as well as the Cronbach’s alpha reliability of the EFL reading comprehension test and appraisal confidence. The skewness and kurtosis statistics for EFL reading comprehension test were within the range of ±1, suggesting that the data were strictly normally distributed. The reliability (Cronbach’s alpha) for the test ($\alpha=0.75$) showed that the test was reliable for the given participants. Similarly, the skewness and kurtosis statistics for appraisal confidence rating scales were within the range of ±3 indicating that the data were generally normally distributed. Cronbach’s alpha coefficient for the confidence rating scales for the whole test was good ($\alpha = 0.88$). In this research, the raw score of the appraisal confidence rating scales were used in the data analysis to answer the research questions.

Table 2: Descriptive statistics and reliability of the EFL reading comprehension test and appraisal confidence (N = 203)

	Min.	Max.	Mean	Std. Deviation	Skewness	Kurtosis	Cronbach’s alpha
EFL reading comprehension test	19.44	94.44	52.35	14.18	-0.07	-0.08	0.75
Appraisal confidence	46.81	99.43	86.84	8.28	-1.09	-2.35	0.88

ANALYSIS TO ANSWER THE RESEARCH QUESTION

- What is the nature of EFL learners' appraisal confidence and appraisal calibration for the reading comprehension test?

As discussed in the method section, the test scores were converted into percentages to be in parallel with the appraisal confidence ratings. Table 3 presents the descriptive statistics of students reading performance scores on the EFL reading comprehension test. Despite the high observed maximum scores on EFL reading comprehension test (i.e., 94.44), the test mean scores indicated that this group of EFL learners did not perform the test well (mean score 52.35 Percent). With respect to variability, the standard deviation was 14.18. As seen in Table 3, the examinees' average appraisal confidence score was 86.84, showing that their average appraisal confidence in their performance was actually higher than their average test performance (i.e., 52.35).

Table 3: Descriptive statistics of test performance and appraisal confidence (N=203)

	Minimum	Maximum	Mean	Std. Deviation
Overall EFL reading performance	19.44	94.44	52.35	14.18
Appraisal confidence	46.81	99.43	86.84	8.28

The EFL learners' appraisal calibration scores (i.e., calibration score +34.49) obtained by subtracting test performance in percentage terms from appraisal confidence ratings indicated that the learners were highly overconfident. Figure 2 shows the mean appraisal calibration of the EFL learners.

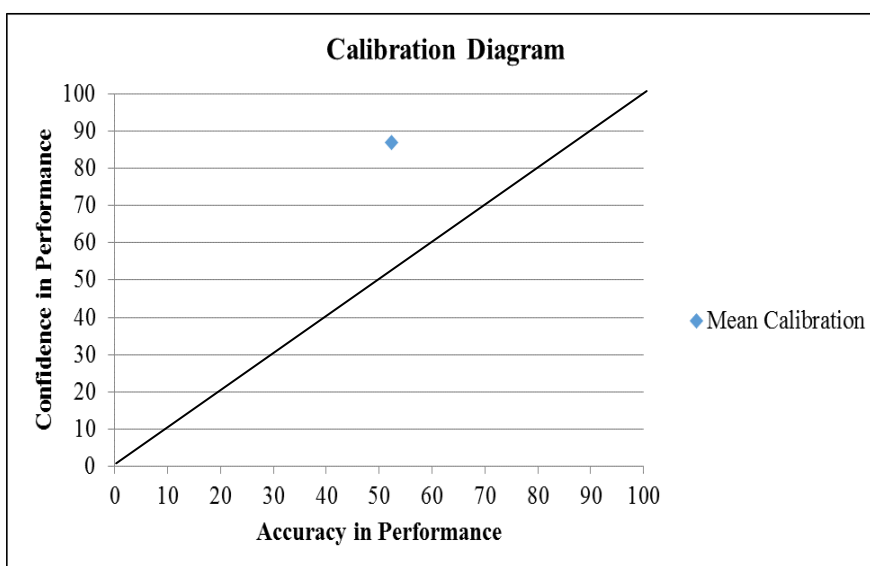


Figure 2: Mean appraisal calibration diagram of students

However, the appraisal calibration diagram of individual EFL learners based on the whole test indicates that the majority of them whose scores were at a 20–40% success rate were highly overconfident (the dots above the unity line), compared to those who performed much better (e.g., at the 70% success rate). Figure 3 shows appraisal calibration of all EFL learners based on the whole test (N = 203).

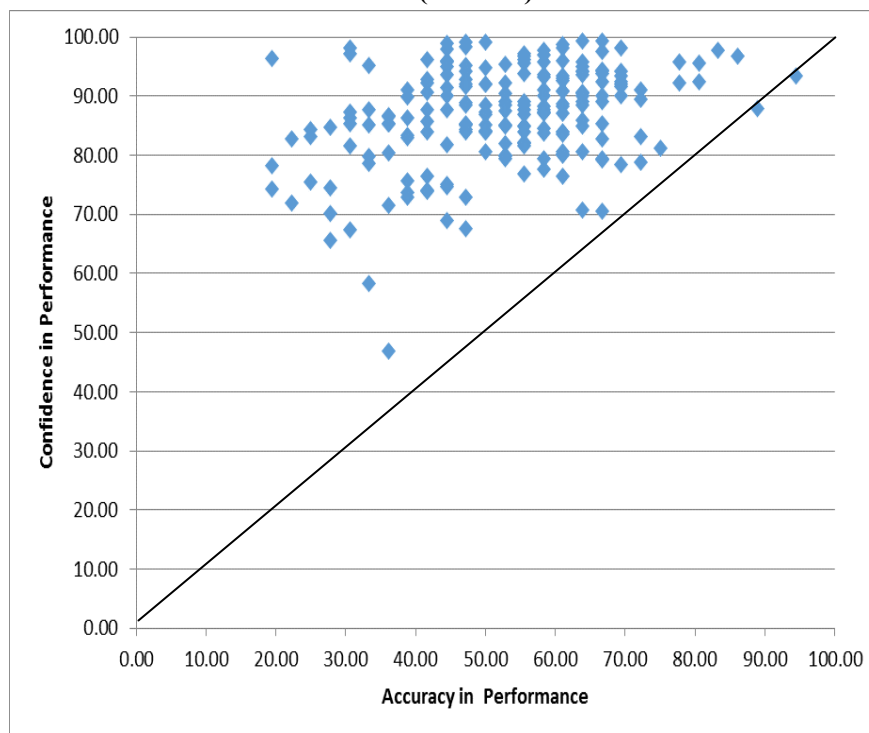


Figure 3: Appraisal calibration diagram of all EFL learners (N = 203)

CONCLUSIONS AND IMPLICATIONS

Reading comprehension is a complex process of making meaning of reading texts which is influenced not only by linguistic factors but also by metacognitive factors (such as comprehension monitoring and appraisal confidence). If learners are not realistic about what they know and what they do not know, they are unlikely to bring improvement in their reading skills. The research has shown that the EFL learners were highly overconfident. As it can have serious consequences, it is necessary to reduce learners’ overconfidence. Once the teachers know the nature of learners’ calibration, they can use appropriate strategies to make them realistic about their performance and develop appropriate study habit. As previous research has indicated, teachers can provide feedback to learners, ask them to provide justification for their answers or provide incentives to

improve their calibration (Arkes *et al.*, 1987; Hacker *et al.*, 2009) so as to help them bring improvement in their reading.

LIMITATIONS AND IMPLICATIONS FOR FURTHER RESEARCH

Because no single study can be perfect, it is worthwhile to note some key limitations of this study and to discuss how future research may consider improvements in its design. Although the study has begun to unlock the nature of appraisal confidence and calibration, the findings were skewed not only by the instruments used (the test and the confidence rating scales) but also by the characteristics of the EFL learners (e.g., by their motivation to do well in the test; their levels of English proficiency). Further research this area should consider involving learners having different levels of language proficiency to better understand the nature as well as role of appraisal calibration in improving EFL learners' reading comprehension.

WORKS CITED

- Akhondi, M., Malayeri, F. A., & Samad, A. A. (2011). How to Teach Expository Text Structure to Facilitate Reading Comprehension. *The Reading Teacher*, 64(5), 368–372. <http://doi.org/10.1598/RT.64.5.9>
- Alderson, J. C. (2002). *Assessing Reading*. Cambridge: Cambridge University Press.
- Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two Methods of Reducing Overconfidence. *Organizational Behavior and Human Decision Processes*, 39(1), 133–144. [http://doi.org/10.1016/0749-5978\(87\)90049-5](http://doi.org/10.1016/0749-5978(87)90049-5)
- Baker, L. (1979). Comprehension Monitoring: Identifying and Coping with Text Confusions. *Journal of Reading Behaviour*, 11(4), 365–374.
- Berne, J. (2004). Think-aloud Protocol and Adult Learners. *Adult Basic Education*, 14(3), 153–173.
- Bjorkman, M. (1992). Knowledge, Calibration, and Resolution: A linear Model. *Organizational Behavior and Human Decision Processes*, 51, 1–21.
- Block, B. L. (1992). See how they Read: Comprehension Monitoring of L1 and L2 Readers. *TESOL Quarterly*, 26(2), 319–343.
- Dinsmore, D. L., & Parkinson, M. M. (2013). What are Confidence Judgments Made of? Students' Explanations for their Confidence Ratings and what that Means for Calibration. *Learning and Instruction*, 24, 4–14. <http://doi.org/10.1016/j.learninstruc.2012.06.001>

- Fry, E. (1977). Fry's Readability Graph: Clarifications, Validity, and Extension to Level 17, *21*(3), 242–252.
- Grabe, W., & Stoller, F. L. (2011). *Teaching and Researching Reading* (2nd ed). Harlow, England: Longman/Pearson.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (2009). *Handbook of Metacognition in Education*. New York: Routledge.
- Hadwin, A. F., & Webster, E. A. (2013). Calibration in Goal Setting: Examining the Nature of Judgments of Confidence. *Learning and Instruction, 24*, 37–47. <http://doi.org/10.1016/j.learninstruc.2012.10.001>
- Han, F. (2012). Comprehension Monitoring in Reading English as a Foreign Language. *New Zealand Studies in Applied Linguistics, 18*(1), 36–49.
- Jonsson, A.-C., & Allwood, C. M. (2003). Stability and Variability in the Realism of Confidence Judgments Over time, Content Domain, and Gender. *Personality and Individual Differences, 34*, 559–574.
- Kleitman, S., & Moscrop, T. (2010). Self-confidence and Academic Achievements in Primary-school Children: Their Relationships and Links to Parental Bonds, Intelligence, Age, and Gender. In *Trends and Prospects in Metacognition Research* (pp. 293–326). New York: Springer.
- Kleitman, S., & Stankov, L. (2007). Self-confidence and Metacognitive Processes. *Learning and Individual Differences, 17*(2), 161–173.
- Kumar, K. B., & Welsun, D. van. (2013). *Knowledge-based Economies and Basing Economies on Knowledge Skills: A Missing Link in GCC Countries*. Washington: RAND Corporation.
- Maclellan, E. (2014). How Might Teachers Enable Learner Self-confidence? A Review Study. *Educational Review, 66*(1), 59–74.
- Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of College Intermediate French Students: Relation to Achievement and Motivation. *Language Learning, 57*(3), 417–442.
- Nuttall, C. E. (2005). *Teaching Reading Skills in a Foreign Language*. Oxford: Macmillan Education.
- Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., & Roberts, R. D. (2002). The Role of Individual Differences in the Accuracy of Confidence Judgments. *The Journal of General Psychology, 129*(3), 257–299.
- Phakiti, A. (2016). Test Takers' Performance Appraisals, Appraisal Calibration, and Cognitive and Metacognitive Strategy Use. *Language Assessment Quarterly, 13*(2), 75–108.

- Richardson, W. (2010). *Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms* (3rd ed). Thousand Oaks, Calif: Corwin.
- Stankov, L., & Lee, J. (2008). Confidence and Cognitive Test Performance. *Journal of Educational Psychology*, 100(4), 961–976. <http://doi.org/10.1037/a0012546>
- Stankov, L., Pallier, G., Danthiir, V., & Morony, S. (2012). Perceptual Underconfidence: A Conceptual Illusion? *European Journal of Psychological Assessment*, 28(3), 190–200.
- Van Doren, C. L. (1992). *A History of Knowledge: Past, Present, and Future*. New York, N.Y.: Ballantine Books.
- van Loon, M. H., de Bruin, A. B. H., van Gog, T., & van Merriënboer, J. J. G. (2013). Activation of Inaccurate Prior Knowledge Affects Primary-school Students' Metacognitive Judgments and Calibration. *Learning and Instruction*, 24, 15–25.