

Do Learners Know 'What they Know' in EFL Reading?

Madhu Neupane Bastola

Abstract

This research examines how realistic EFL learners are in making judgement of their EFL reading comprehension performance. Appraisal confidence and appraisal calibration have been used as frameworks in this study. Appraisal confidence refers to the degree to which learners believe that their answer to a test item is correct or appropriate. It is expressed in percentage terms (e.g., 0%, 25%, 50%, 75%, or 100%). Appraisal calibration examines the accuracy of learners' appraisal confidence. Being realistic about accuracy in performance (i.e. appraisal calibration) is considered to be an indicator of effective learning. However, despite its importance, appraisal confidence and appraisal calibration has gained little attention especially in EFL reading. In order to minimize this knowledge gap, this study examined the appraisal confidence and appraisal calibration of eighty-five EFL students studying Master of Education (M. Ed) with specialization in English in Tribhuvan University. An EFL reading comprehension test specifically designed for the study and appraisal confidence rating scales incorporated in the same EFL reading comprehension test were used as the tools for data collection. The findings of the study show that the students in general were not well calibrated. However, high performers were better calibrated than the low performers. The implications of the study for teaching and suggestions for further research are discussed.

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

Introduction

In the present day knowledge based economy, knowledge plays a key role in academic, professional as well as economic success (see Kumar & Welsum, 2013). In this context, the role of reading as a source of information and knowledge cannot be overstated as modern citizens are unlikely to succeed if they are not skilled readers (Grabe, 2009). In the same way, the need for being a skillful EFL reader has been a much desired goal for many especially in EFL context. As English is a global language, 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 (see Grabe, 2009; Grabe & Stoller, 2011). Furthermore, it is an avenue for EFL learners to gain authentic exposure so as to increase their English language proficiency (see Grabe & Stoller, 2011; Richards & Renandya, 2013). However, for an alarmingly large number of adult EFL learners, reading complicated texts is a daunting task, a great barrier for their success (Berne, 2004). For this reason, the pertinent question is: How can EFL learners' reading abilities can be improved?

For answering the question, first, we need to know what it means to be a skillful reader or to be successful in reading. Successful reading comprehension is usually defined as the reader's understanding of the message expressed by the writer (see Nuttall, 1996). Such understanding requires speed, efficiency, purpose, and constant interaction with the text on the part of readers (see Grabe, 2009; Grabe & Stoller, 2011). It also necessitates 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 organization of the text, interpreting the text according to the situation and context, using background knowledge, and making inferences). Though readers, through extensive practice, can develop automaticity in executing lower level processes in reading, their metacognition- knowledge about cognition-plays a significant role in the execution of higher level processes (Block, 1992; Casanave, 1988; Grabe & Stoller, 2011; Mills, Pajares, & Herron, 2007). One of the components of metacognition that is shown to be positively associated with reading comprehension is learners judgement about what they can do and what they cannot do as readers (e.g Kleitman & Moscrop, 2010) and appraisal calibration (Phakiti, 2016). Therefore, in this study an attempt has been made to examine learners' appraisal confidence and appraisal calibration in EFL reading comprehension performance.

Appraisal confidence and its measurement

Appraisal confidence is defined as learners' subjective expression of decisions about the likely outcomes of their performance in a test (see Harvey, 1997; Jonsson & Allwood, 2003; Kleitman & Moscrop, 2010; Phakiti, 2016; Stankov, Pallier, Danthair, & Morony, 2012). To examine appraisal confidence, learners are asked to report the accuracy of their performance immediately after they

provide a response to a test item (see Kleitman & Moscrop, 2010). For example, they may be asked to indicate their belief that a response to a test item is correct with a 25%, 50%, 75% or 100% probability (see Phakiti, 2016) as shown in the following example:

1. *According to the text, American dream refers to....*
 - a. *going to America for better life*
 - b. *having access to computer and the Internet*
 - c. *having a house, a car, and modern electrical appliances for having a better living standard*
 - d. *having per capita income equal to that of American people*

25%	50%	75%	100%
-----	-----	-----	------

Previous research shows that people usually overstate their knowledge which may be because of overconfidence effect (i.e. confidence judgement that is higher than accuracy in performance) or hard-easy effect (i.e. increase in overconfidence with the increase in the difficulty of question) (Gigerenzer, Hoffrage, & Kleinbölting, 1991). To explain the effect of spontaneous confidence, Gigerenzer et al. (1991) have proposed two theories: local mental model (LMM) and probabilistic mental model (PMM). They claim that learners first use LMM to solve the given problem. In doing so, they retrieve information from memory, and decide the most appropriate option by eliminating the options that they know are wrong. If these strategies work, learners can produce the correct answer with 100% confidence. However, sometimes memory fails and what is considered to be correct with 100% certainty may be wrong indeed. If

this happens, learners may come up with a wrong answer with a greater degree of certainty.

When LMM is not sufficient to come to an acceptable answer, PMM is constructed to answer the question. PMM is based on inductive inferences. According to Gigerenzer et al. (1991), to make inductive inferences various sources of information such as reference class (that is the class of reference that an answer belongs to) and probability cues. However, in such cases, as answers are based on inferences, uncertainty resulting in under or overconfidence is the outcome.

Phakiti (2016) has classified appraisal confidence in two types: single-case appraisal confidence and relative frequency appraisal confidence. The former refers to appraisal confidence related to a single test item whereas the latter refers to appraisal confidence associated with the whole test. In this research, the relative frequency appraisal confidence (i.e. the appraisal confidence associate with a test as a whole) has been examined. Appraisal confidence in a test as a whole refers to the average score for all the attempted test items (see Kleitman, Mark, Young, Lau, & Livesey, 2011).

Appraisal calibration and its measurement

Appraisal calibration, which is based on appraisal confidence, refers to the match between the estimates of correctness and the accuracy of answers (see Harvey, 1997; Jonsson & Allwood, 2003; Kleitman & Moscrop, 2010). It expresses the correspondence between subjective and objective probability (Bjorkman, 1992). In calibration research, it is common to compare the mean scores of appraisal calibration and test success (Phakiti, 2016).

The same procedure has been followed in this research.

Computing calibration is fairly straightforward. First, both overall test scores and appraisal confidence in the overall test are converted into percentage. Then the overall performance in a test is subtracted from overall scores of appraisal confidence. On the basis of this calculation method, test takers are said to be well calibrated when their appraisal confidence level matches to 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 a learner's reported appraisal confidence is 75% and they secure 75% in the given test, the test taker is considered to be well calibrated ($75\% - 75\% = 0$). In contrast, the mismatch between the accuracy of judgement and objective accuracy is called miscalibration (Maclellan, 2014; Phakiti, 2016; Stankov et al., 2012). For example, if the appraisal confidence is 75%, but the test performance is 50%, the learner is said to be overconfident ($75\% - 50\% = 25\%$). In the same way, if the appraisal confidence is 50%, but the test performance is 70%, the learner is said to be underconfident ($50\% - 75\% = -25\%$). As exemplified above, zero, positive, and negative calibration scores show perfect calibration, overconfidence, and underconfidence respectively. Research has shown that people are typically overconfident in the judgement they express about their performance (Arkes, Christensen, Lai, & Blumer, 1987). A group calibration diagram can be used to present test takers' appraisal calibration vividly.

Figure 1 presents an example of an appraisal calibration diagram in which the 45° line (called a unity line) represents the test performance. Calibration scores that fall on the line represent the learners who are perfectly calibrated. In contrast, the

calibration scores that fall above or below the unity line represent the learners who are overconfident and underconfident respectively. Though an ideal appraisal calibration score should be on or close to the unity line, a learner is considered to be well-calibrated when the appraisal calibration score is within $\pm 5\%$ (Phakiti, 2016) or $\pm 10\%$ (Stankov & Lee, 2008).

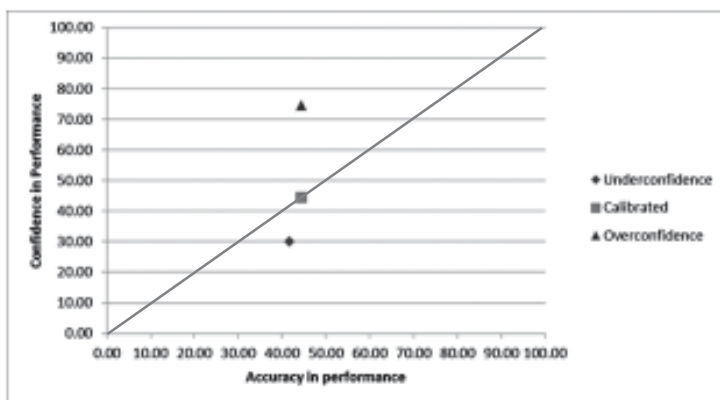


Figure 1. Appraisal calibration diagram for an individual (adopted from Neupane, 2016, p. 100)

Review of related literature

Appraisal confidence has been widely researched in the area of psychology and is shown to be a component of metacognitive self-monitoring (Kleitman & Moscrop, 2010; Kleitman & Stankov, 2007). There is a body of research showing positive correlation between appraisal confidence and L2 proficiency (e.g. Clement, Dörnyei, & Noels, 1994; Stankov & Crawford, 1997). Clement et al.'s (1994) study on 301 secondary school children in Budapest concluded that self-confidence influences L2 proficiency both directly and indirectly through learners' attitude towards and efforts expended on learning English. Similarly, Stankov and Crawford's (1997) investigation of individual differences in confidence judgements showed significant positive correlation between vocabulary test score and confidence rating ($r = 0.63$)

though the learners' were somehow overconfident. Similar positive correlation between accuracy and appraisal confidence ($r = 0.43$) was reported in Jonsson and Allwood (2003). Unlike in Stankov and Crawford's (1997) study, secondary school students' ($N = 79$) in Jonsson and Allwood's (2003) study were found to be better calibrated.

Similarly, Stankov and Lee (2008) assessed confidence of 824 native speakers during the administration of reading and listening sections of Testing English as a Foreign Language Internet Based Test (TOEFL iBT). The participants took all the components of TOEFL iBT in the morning and participated in the selected versions of reading and listening along with

confidence rating during the day in the interval of four hours. The result showed somewhat stronger associations between reading accuracy and appraisal confidence scores (ranging from 0.469 to 0.605) than between listening accuracy and appraisal confidence scores (ranging from 0.358 to 0.490). Similarly, groups with lower performance accuracy showed higher overconfidence. More recently, Stankov et al.'s (2012) large scale study ($N = 1,786$) examined the relationship between appraisal confidence and accuracy scores, prediction of achievement scores from appraisal confidence and self-beliefs of teenage secondary three (equivalent to Grade 9 in the USA) students in Singapore. Measures of self-belief questionnaires, achievement tests in English and appraisal confidence rating scales were administered online by using Qualtrics.Inc. The study showed a positive correlation between appraisal confidence and test scores ($r = 0.56$), and the confidence to be a better self-belief predictor of achievement among the

other self-belief factors (self-efficacy, anxiety, and self-concept) included in the study.

Like appraisal confidence, appraisal calibration has widely been researched (see, Dinsmore & Parkinson, 2013) especially in the field of psychology. Some 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 which aimed at reducing the overconfidence of undergraduate students showed that by providing feedback on learners' performance (answer they provide to individual questions) and asking them to provide justification for the answer they provide, learners' calibration can be improved. However, Hacker et al.'s (2009) quasi-experimental study involving 137 college students showed no significant effect of asking learners to reflect on their confidence judgement and providing incentives in improving learners' calibration. The study showed that the learners' appraisal calibration differed significantly between higher-performers and low performers. They concluded that by increasing the subject knowledge of the learners, calibration can be improved.

Further research on calibration 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) (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 the learners who were performing better at university tended to be better calibrated. Previous 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 showed 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 and their appraisal confidence could predict just above one third of the test performance variance. Similarly, they tended to be underconfident in easy questions but overconfident in difficult questions and their appraisal calibration was not strongly related to reported metacognitive strategy use. Similarly, Neupane's (2016) study on master level students (N = 203) calibration on EFL

reading showed that they poorly calibrated as they were highly overconfident.

The review of literature shows that despite the abundance of research in appraisal confidence and appraisal calibration in the field of psychology, these aspects have got little attention in EFL reading. This lack of research warrants further research in these aspects.

Research questions

The present study has addressed the following research questions:

1. What is the nature of learners' appraisal confidence and appraisal calibration in EFL reading comprehension performance?
2. Do learners at different levels of reading comprehension differ in terms of appraisal calibration?

Methodology

Setting and participants

This research was carried out at the Department of English Education, Tribhuvan University, Kathmandu. The participants of the study were the students studying for the Master of Education (M.Ed.) degree with specialization in English. About 100 students took part in the study but usable data came from just 85 students due to incomplete data. Out of 85 students, 46 (54.1%) were male and 39 (45.9%) were female.

Research instruments

In order to answer the research questions, an EFL reading comprehension test and appraisal confidence scales incorporated in the reading comprehension test were used.

EFL reading comprehension test

For the purpose of this research, a reading comprehension test was prepared. Reading comprehension comprises a number of skills and strategies (see Alderson, 2000; Grabe & Stoller, 2011; Nuttall, 1996). For this research, the specifications of reading skills were prepared. Six different skills namely *identifying factual information*, *making inferences*, *getting meaning of words in a context*, *identifying main ideas*, *identifying purpose and attitude of the writer*, and *identifying references* were focused in the study.

An expository text - informational texts usually written in the present tense and containing a high number of technical words (Akhondi, Malayeri, & Samad, 2011)- was selected from Friedman (2008) for the purpose of the study. The rationale for selecting an expository text was that university students are required to read high number of such texts across the courses they take. The texts had the difficulty level of grade 13+, the basic level for university students, according to Fry's (1977) readability formula. In order to test the skills specified above, 20 multiple choice questions (each worth one mark) based on the selected reading text were designed.

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). Since all the multiple choice questions in the EFL reading test used in the current research had four alternatives, a four-point relative frequency appraisal confidence scales (i.e., 25%, 50%, 75% and 100%) were used. The relative frequency appraisal confidence scales were embedded into each test

question. The questions were designed to allow the learners to record both their answers and appraisal confidence estimates. The learners were instructed to rate their appraisal confidence immediately after they answered each question. Given the additional cognitive load that the students had in evaluating their confidence about the answer they had given, the learners were given one-hour time to complete the exam.

Data preparation and analysis

To prepare the data collected from a reading comprehension test and appraisal confidence rating scales for analysis, three main steps were taken. First the data were entered into SPSS version 22 for PC. Overall scores for EFL reading comprehension test as well as appraisal confidence were converted into percentage. At item level each test score was dichotomous (0 or 1). However, at the test level as a whole the scores were continuous due to addition of the scores of series of questions (Phakiti, 2016). Appraisal confidence ratings were recorded at 25%, 50%, 75%, and 100%. At an item level, appraisal confidence was ordinal, but at a test level as a whole, appraisal confidence scores were also continuous because they were aggregated across various appraisal confidence ratings in a series of test questions. In SPSS spreadsheet, the score for a test item and its appraisal confidence were paired in the data entry. The descriptive statistics of each data set were first computed to examine central tendencies, variability, and distribution of raw data score. The reliability and internal consistency of the research instruments (i.e. the EFL reading comprehension test and appraisal confidence rating scales) were calculated by using Cronbach's alpha coefficient.

To address research question 1 (the nature of learners' appraisal confidence and appraisal calibration in EFL reading comprehension), the mean percentage scores of the learners' EFL reading comprehension performance and appraisal confidence were computed for the whole test. Calibration scores were also computed to examine the extent to which learners were realistic in their appraisal confidence. Appraisal calibration diagrams were created by using Microsoft Excel. Pearson correlations were then computed to examine the relationship between test performance and appraisal confidence for the whole test. Similarly, in order to answer research question 2 (the differences between learners having different levels of reading comprehension performance in terms of appraisal calibration), a one-way analysis of variance (hereafter one-way ANOVA) with post hoc analysis (Bonferroni) was used as it is more robust than an independent sample t-test (Field, 2009; Phakiti, 2014).

Results and discussion

First the preliminary analysis of the research instruments was carried out. Table 1 presents the descriptive statistics as well as the Cronbach's alpha reliability of the EFL reading comprehension test and appraisal confidence rating scales. In reliability analysis four test items (q3, q7, q9, & q17) contributing negatively for the test reliability were deleted and only 16 items were submitted for further analysis. After removing the four test items, the reliability (Cronbach's alpha) of the EFL reading comprehension test was 0.60. It showed that the test was moderately reliable for the given participants. The moderate reliability estimate can be explained by the restricted range of variability in the students' performance (Phakiti, 2016). However, the reliability of the appraisal confidence rating scales was

very good (i.e. $\alpha = 0.90$). The skewness and kurtosis statistics for EFL reading comprehension test and appraisal confidence were within the range of ± 1 indicating that the data were generally normally distributed.

Table 1: Descriptive statistics and reliability of the EFL reading comprehension test and appraisal confidence (N = 85)

	Min.	Max.	Mean	Std. Deviation	Skewness	Kurtosis	Cronbach's alpha
EFL reading comprehension test	25.00	100.00	66.39	16.92	-0.25	0.59	0.70
Appraisal confidence	42.19	100.00	82.58	13.80	-1.071	0.88	0.90

After the analysis of instruments, data were analyzed to answer the research questions raised in the study.

Research Question 1: What is the nature of EFL learners' appraisal confidence and appraisal calibration for the EFL reading comprehension test?

As discussed in the method section, the test scores and appraisal confidence scores were converted into percentage so as to make them parallel. Table 2 presents the descriptive statistics of students' scores on the EFL reading comprehension test. Despite the high observed maximum scores on EFL reading comprehension test (i.e. 100%), the test mean scores indicated that the performance of the students who participated in the test was moderate (mean score 66.39%). With respect to variability, the standard deviation was 16.92 and the scores ranged from 25% to 100%. As seen in Table 2, the students' appraisal confidence score ranged from 42.19% to 100% with the mean score of 82.58. This indicates that the students' average appraisal confidence score (i.e. 82.58) was actually higher than their

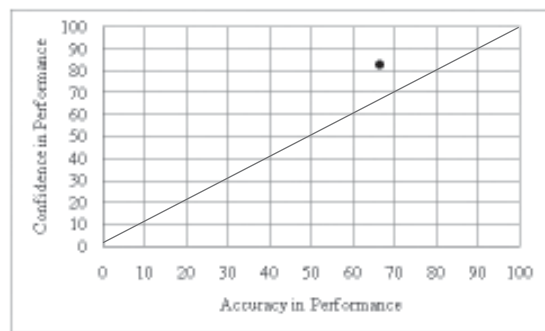
average performance (i.e., 66.39%) in the EFL reading comprehension test. Learners' appraisal calibration scores (obtained by subtracting test performance scores from appraisal confidence ratings in percentage terms) shows that they were overconfident in the EFL reading comprehension performance as the average calibration score was +16.19%. Figure 2 shows the group appraisal calibration graph of students in the overall test.

In the Figure the dot representing the mean appraisal confidence score is above the unity line showing that they were overconfident in judging their performance.

Table 2: Descriptive statistics of EFL reading comprehension performance, appraisal confidence, and appraisal calibration (N=85)

	Min.	Max.	Mean	SD	Skewness	Kurtosis
EFL reading comprehension test	25.00	100.00	66.39	16.92	-0.25	0.59
Appraisal confidence	42.19	100.00	82.58	13.80	-1.071	0.88
Appraisal calibration	-17.19	60.94	16.19	19.00	0.25	-0.64

Figure 2. Group appraisal calibration diagram (N = 85)



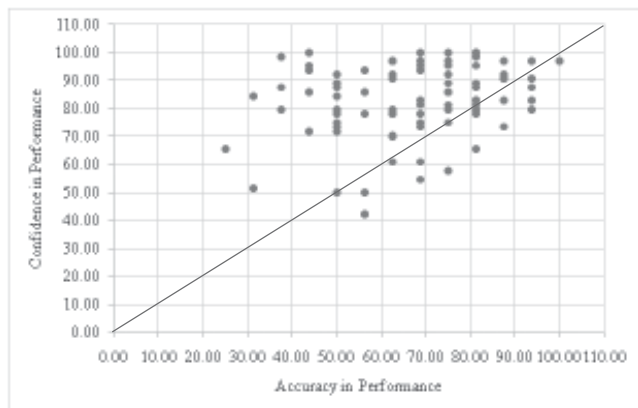


Figure 3. Appraisal calibration diagram of all students on the whole test (N = 85)

Correlational analysis between EFL reading comprehension performance and appraisal confidence

The correlation coefficient between learners’ appraisal confidence and actual performance can provide some indication about the calibration of the group (Phakiti, 2016). Table 3 shows the correlation between learners’ EFL reading comprehension performance and appraisal confidence.

Table 3: Correlations between EFL reading comprehension performance and appraisal confidence (N = 85)

	<i>Appraisal confidence</i>
EFL reading comprehension performance	0.25*

* Correlation is significant at the 0.05 level (2-tailed).

As seen in Table 3, the correlation between EFL reading comprehension performance and appraisal confidence was 0.25 ($R^2 = 0.06$). The correlation was significant at 0.05 level (i.e. $p < 0.05$). This

statistically significant relationship implies that the students who performed better in the test tended to rate their appraisal confidence more highly. However, the correlation coefficient suggests that only 6% of reading comprehension performance in the test was explained by appraisal confidence. This shows poor calibration of the learners as correlation of 0.70 or above is considered to be an indicator of good calibration (Phakiti, 2016). In conclusion much

lower correlation (i.e., $r = 0.25$) observed in the study suggests that the students were poorly calibrated. The findings are similar to those in the previous study (e.g. Hacker, Bol, & Bahbahani, 2008; Neupane, 2016; Phakiti, 2016).

Research question 2: How do learners at different levels of EFL reading comprehension performance differ in terms of calibration?

As discussed in methodology section, a one-way ANOVA was examined to investigate group differences in calibration in terms of their levels of performance in EFL reading comprehension test: high performers (70% and above), moderate performers (50 to 69%) and low performers (below 50%). Table 4 presents the summary of descriptive statistics different groups.

Table 4: Descriptive statistics of group differences in appraisal calibration

	N	Mean	SD	Min.	Max.	Skewness	Kurtosis
High Performers	35	6.12	13.52	-17.19	53.13	1.04	3.06
Moderate Performers	40	18.00	16.92	-14.06	42.19	-0.59	-0.77
Low Performers	10	44.22	12.46	20.31	60.94	-0.77	0.18
Total	85	16.19	19.00	-17.19	60.94	0.25	-0.64

Table 4 shows that students with a high level of EFL reading comprehension were better calibrated (mean score 6.12%) than

the moderate and low performers. To find out whether the differences among groups were significant, a one-way ANOVA was conducted. The skewness and kurtosis statistics were within the range of ± 1 to ± 3 suggesting that the group data for appraisal calibration was normally distributed. Table 5 shows the ANOVA of group differences.

Table 4: ANOVA of group differences

	Sum of Squares	df	Mean Square	F ratio	Sig.	Eta squared (ζ^2)
Between Groups	11539.13	2.00	5769.56	25.20	0.00	0.38
Within Groups	18773.39	82.00	228.94			
Total	30312.52	84.00				

According to Table 4, there were statistically significant differences between three groups of learners in terms of appraisal calibration ($F [2, 84] = 25.20, p < 0.01, \zeta^2 = 0.38$). The eta squared of 0.38 shows 38% of differences in calibration can be explained by level of EFL reading comprehension performance. However, the ANOVA does not explain where the statistically significant differences occurred among the groups. To explain this, a Bonferroni *post hoc* test was performed. Table 5 presents the result of *post hoc* comparison.

Table 5: Bonferroni post hoc test for multiple comparisons of groups

	(I) Level of Comprehension	(J) Level of Comprehension	Mean Difference (I-J)	Sig.	Cohen's d	Percentage of Nonoverlap
High Performers	Moderate Performers		-11.88*	0.00	0.77	43%
	Low Performers		-38.10*	0.00	2.93	81.1%
Moderate Performers	Low Performers		-26.21*	0.00	0.37	21.3%

*The mean difference is significant at the 0.05 level.

As can be observed from result of the post hoc test in Table 5, statistically significant differences were observed between all levels of learners ($p < 0.01$) in appraisal calibration. There was 43% difference between the mean calibration of high performers and moderate performers ($p < 0.01, \text{Cohen's } d = 0.77$) and more than 81% difference between high and low performers ($p < 0.01, \text{Cohen's } d = 2.93$) (see <http://www.uccs.edu/lbecker/effect-size.html> for the interpretation of Cohen's d). In conclusion, the one-way ANOVA shows that high performers (securing 70% and above) were better calibrated than the moderate performers (securing 50-69%) and low performers (securing below 50%) in EFL reading comprehension performance. In other words, it shows that EFL reading comprehension performance had significant effect on appraisal calibration. The findings corroborate to previous research (e.g., Hacker et al., 2008).

Conclusions and implications

As discussed in the previous section reading comprehension is complex in nature. If reading comprehension is taken to be analogous to driving a car to reach a destination, lower level processes constitute the fuel and engine, while higher level processes refer to driving (Grabe & Stoller, 2011). Appraisal confidence and appraisal calibration, which affect higher level processes, in reading show how good learners are in judging their performance. This research has shown that the learners in general were not realistic in making such judgement as they tended to

overstate their performance. In other words, they were not well-calibrated. One-way ANOVA analysis showed that high performers were significantly better calibrated than moderate and low performers. To put it differently, low performers thought they knew more than they actually knew.

Without being aware of what they know and what they do not know, EFL learners may not be able to bring improvement in their reading. As lack of calibration may have serious consequences, it is incumbent on teachers to promote learners' calibration. As previous research has indicated, teachers can provide feedback to learners and ask them to provide justification for their answers (Arkes et al., 1987) so as to help them bring improvement in their reading. They may also like to incorporate appraisal confidence rating scales in formative assessment (Kleitman & Stankov, 2007).

The present study had some limitations which future research may consider to address. First, all the participants in the study belonged to the same department. Incorporating large sample randomly selected from different contexts may improve external validity of the findings. Second, only single expository text was used in this research. As learners are required to read varied types of text, incorporating different types of texts along with different types of questions would be worth considering.

References

- 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. <https://doi.org/10.1598/RT.64.5.9>
- Alderson, J. C. (2000). *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. [https://doi.org/10.1016/0749-5978\(87\)90049-5](https://doi.org/10.1016/0749-5978(87)90049-5)
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, N.J: Prentice-Hall.
- 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.
- Casanave, C. P. (1988). Comprehension monitoring in ESL reading: A neglected essential. *TESOL Quarterly, 22*(2), 283-302.
- Clement, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence and group cohesion in the foreign language classroom. *Language Learning, 44*(3), 417-448.
- 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. <https://doi.org/10.1016/j.learninstruc.2012.06.001>

- Friedman, T. L. (2008). *Hot, flat, and crowded: Why we need a green revolution, and how it can renew America* (1st ed). New York: Farrar, Straus and Giroux.
- Fry, E. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading*, 21(3), 242-252.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98(4), 506-528. <https://doi.org/10.1037/0033-295X.98.4.506>
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York: Cambridge University Press.
- Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed). Harlow, England: Longman.
- Hacker, D. J., Bol, L., & Bahbahani, K. (2008). Explaining calibration accuracy in classroom contexts: The effects of incentives, reflection, and explanatory style. *Metacognition and Learning*, 3(2), 101-121. <https://doi.org/10.1007/s11409-008-9021-5>
- 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. <https://doi.org/10.1016/j.learninstruc.2012.10.001>
- Harvey, N. (1997). Confidence in judgement. *Trends in Cognitive Sciences*, 1(2), 78-82.
- 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., Mark, K., Young, S., Lau, P., & Livesey, D. (2011). *Something about metacognition: Self-confidence factor(s) in school aged children*. New York: Nova Science Publishers, Inc.
- 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 A. Efklides, & P. Misailidi (Eds.), *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. <https://doi.org/10.1016/j.lindif.2007.03.004>
- Kumar, K. B., & Welsum, D. van. (2013). Knowledge-based economies and basing economies on knowledge skills: A missing link in GCC countries. Washington: RAND Corporation RAND Corporation.
- Maclellan, E. (2014). How might teachers enable learner self-confidence? A review study. *Educational Review*, 66(1), 59-74. <https://doi.org/10.1080/00131911.2013.768601>
- 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.

- Neupane, M. (2016). Calibration in EFL reading: Examining the nature of judgement of confidence. *Tribhuvan University Journal*, 29(1), 97-110.
- Nuttall, C. (1996). *Teaching reading skills in a foreign language*. Oxford: Heinemann.
- Phakiti, A. (2016). Test takers' performance appraisals, appraisal calibration, and cognitive and metacognitive strategy use. *Language Assessment Quarterly*, 13(2), 75-108. <https://doi.org/10.1080/15434303.2016.1154555>
- Richards, J. C., & Renandya, W. A. (Eds.). (2013). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.
- Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence*, 25(2), 93-109.
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100(4), 961-976. <https://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 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. <https://doi.org/10.1016/j.learninstruc.2012.08.005>

The Author: Madhu Neupane Bastola is a lecturer at the Department of English Education at Tribhuvan University where she lectures in interdisciplinary readings and second language research methodology to postgraduate students. She has completed M.Ed. in ELT and M.A. in English literature from Tribhuvan University, Nepal and M.Ed. (TESOL) from the University of Sydney, Australia. She is interested in second language reading, academic writing, self-assessment, and self-regulated learning.