# Learner Ability Calibration: A Work in Progress

**Todd Leroux**

**Abstract**

*Calibration is the degree of correspondence between a learner's perceived ability to successfully complete a task and actual ability demonstrated. Oral presentations evaluated by explicit criteria were utilized to examine learner ability beliefs assessed by a pretest (learner self-report), then demonstrated and assessed by a post-test (instructor evaluation) based on four criteria. Results suggest learners tended to over-estimate their abilities on eye contact, the use of gestures, and speaking time (length of presentation) with statistical significance, while underestimating their fluency ability though not statistically significant.*

## Introduction

How 'good' are you at rating your ability to accomplish different tasks? For example, how many free-throws from a standard basketball court could you make out of ten tries? Or, how many gyozas could you make in 5 minutes? Or, how many correct answers on a test of second language intermediate sentence structure would you expect to achieve? For most people, the answer to the initial question of "How good are you at…?" is, "Not very." People are not good at rating how much or how well they can accomplish tasks even in domains where they have considerable experience. Further, regarding the above questions, go a little deeper and ask yourself about how you came to the decisions you did? What factors did you weigh, how much did you allot to each and why? Essentially, what you have just done is an exercise in calibration followed by the extremely profitable practice of reflection. Though extensively studied in a variety of domains, little is known about the nature of the decisions (or judgements of confidence) regarding calibration; and, for many the first step is to compare what we think we can do to what we can do based on actual demonstration.

## Literature Review

Calibration is the degree of correspondence between a learner's perceived ability to successfully complete a given task compared to that actually demonstrated (Alexander, 2013). In this discussion 'ability' can refer to acts or actions such as: do, understand, learn, and remember to name a few commonly associated verbs/actions. And, 'given task' is defined as a task that is similar but not completely the same as a proximal, previous task. In other words, some recent experience to draw from is a necessary component in assessing an individual's calibration accuracy (Bol and Hacker, 2001). For example, if a person has never played basketball or thrown a ball into another vessel, then calibrating this person's ability has a high likelihood of error because it is simply based on guessing. The less familiarity one has with a phenomenon, the wilder the guesses. In terms of the actual calibration scores, they are absolute values, or non-directional, which means that over- or under-estimating one's abilities is not important, rather, only the degree of over- or under-estimation is the focus. For directional considerations in calibration accuracy, the term 'bias' is used. Calibration itself is the manifestation of a variety of latent variables or determinants not totally understood; they are ultimately functions of the individual differences that make up all people. Thus, in order to attempt to understand this phenomenon with any kind of structured approached, it is necessary to construct our understanding from a relevant theoretical framework. Self-efficacy Theory (SET) (Bandura, 1977), is an established theory of human motivation. SET is a central tenet of Social Cognitive Theory (SCT), which is the overall framework, and an enduring theory of human development (Bandura, 1986).

Social cognitive theory is a model of causation involving triadic reciprocal determinism. In this model of reciprocal causation, behavior, cognition and other personal factors, and environmental influences all operate as interacting determinants that influence each other bi-directionally (Figure 1).

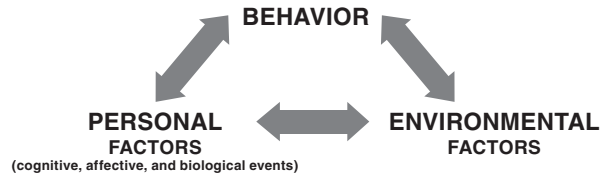


*Figure 1*. Bandura's conception of triadic reciprocal determinism from Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), Annals of child development. Vol. 6. Six theories of child development (pp. 1-60). Greenwich, CT: JAI Press.

SCT was authored by Canadian psychologist Albert Bandura, and based on his 1963 and 1966 'Bobo Dolls' studies, which dramatically illustrated the degree to which behavior is learned by observing others. Social cognitive theory has since been applied with tremendous amounts of success in a wide range of domains (Pajares, 2002).

SET preceded SCT and first emerged in Albert Bandura's game-changing book, Social Learning Theory (1977). Self-efficacy is the belief that one has in his or her ability to exercise control over their level of functioning given environmental demands. Ultimately, it is a theory of the degree that one believes they have personal agency. Personal agency is the ability to control oneself and one's environment as desired or needed to achieve a certain outcome. For example, if a pudgy, middle-aged educator and researcher has a high sense of personal agency and wants to reduce his considerable 'love-handles', then he will be able to control the temptations of buying and consuming (secretly, of course) half-priced, freshly-baked, chocolate cupcakes (with icing and sprinkles), and maintain his healthy, fibre-rich diet. Furthermore, he will adjust his routes and routines to ensure against moments of weakness and diet failure by avoiding the shop sirens calling for his waistline's doom. However, if an individual does not believe that they can control the circumstances of their own existence, then there would be little reason to actively participate or engage tasks of self-interest or self-improvement. Taking the same example of the overweight teacher, if he feels that his physical circumstance is genetic in nature, then it is likely that he feels there is no point in dieting or refusing the half-priced chocolate heavens and will promptly gorge away. In summary, in academic terms, self-efficacy is the belief that an individual has the ability to exercise control over internal and external factors in order to achieve a desired outcome even with faced with a variety of challenges (Bandura, 1996).

As per the above, self-efficacy has broad and powerful implications for individuals as it provides a framework from which to predict future actions that include career and education choices to course-level academic achievement, goal-setting challenge levels, and task performance (Schunk and Pajares, 2009). Research has time and time again evidenced that those with a high sense of self-efficacy are more likely to experience success. They persevere in the face of difficulty, they focus their attention and cognitive resources on finding a solution to a task, they control or regulate internal and/or external conditions that can assist or are detrimental to successful

task completion, and they are able to apply strategies to their advantage. Naturally, these kinds of behaviors are not as evident in those with a low sense of self-efficacy, who tend to give up quickly or engage in task avoidance behaviors. Research has also long since corroborated Bandura's position that self-efficacy comes from 4 sources: mastery experience/previous performance, vicarious learning, social persuasion, and physiological and affective states (Bandura, 1977, 1986, 1997).

Self-efficacy is not to be confused with confidence though they are highly correlated (Wang, et. al., 2014). Confidence, or self-confidence, is a general construct while believed or perceived self-efficacy is domain specific. For example, an athletic individual might feel that would be able to make 8 out 10 free throws with thirty minutes of practice in high confidence, yet they would be much less optimistic about making just 8 gyozas in five minutes with thirty minutes of practice if they have had little experience or bad experiences preparing food items. Thus, in different domains, sports versus cooking, an individual's belief in their ability to succeed would vary, and in many cases, significantly so. If skeptical of this claim, ask almost any language teacher about their confidence level in the teaching of past participles, and then ask them the same question about calculus. One final point of self-efficacy that is relevant to the ensuing coverage of calibration is that it is also culture-specific. A dichotomy: individualistic versus collectivist cultural orientations will have an impact on the triadic reciprocity determinants modeled in SCT (personal, environmental, behavioral). Further, culture-orientations will also impact the degree of strength of each of the four determinants of SET: past experience, vicarious learning, social persuasion, affective/behavioral reactions to stimuli. The United States is an example of an individualistic culture while Japan is an example of a collectivist culture. Thus, for example, social persuasion may have more impact on an individual's thoughts and actions in Japan than on a person in America (Oettingen, 1995). An understanding of SCT and SET is vital in the consideration of calibration and its determinants.

Calibration has been widely studied in many domains and for many years, however, there is limited direct research found for SLA contexts. People, in general, are poorly calibrated as are learners (Bol & Hacker, 2001; Klassen, 2002). It has been commonly measured through 3 procedures: The accuracy index measures the skill for accurate calibration. The C-index can vary from 0 (perfect calibration) to 1 (no calibration at all). The discrimination index (DI) measures the skill to discriminate

between the occurrence and the non-occurrence of an event. The DI ranges from 0 (no discrimination) to 1 (perfect discrimination) and is traditionally calculated as the weighted mean of the squared differences between the mean proportion of correct outcomes in each category and the overall mean proportion of correct outcomes. The O/U-index measures over/under-confidence or students' average tendency to respond with more or with less confidence than their answers to the problems warranted. The index ranges from 0 (complete under-confidence) to 1 (complete overconfidence) and is calculated similarly to the C-index only the differences are not squared (Boekaerts and Rozendaal, 2009).

Research on calibration has shed some light on the determinants in terms of quality calibration (accuracy) in that it is commonly associated with general cognitive ability, level of achievement, or academic ability (in the domain of learning and instruction) and that culture and domain specificity are also at play (Bol, Hacker, Walck, & Nunnery, 2012; Chiu & Klassen, 2010). What is little known is: What is the nature of our calibrations, or according to the literature, what is the nature of our judgements of confidence? Based on work by Dinsmore and Parkinson (2013), an open-ended questionnaire, the development of which was based on Bandura's (1986) model of reciprocal determinism, 4 a priori categories were decided plus an additional 'guessing' category regarding reading calibration. The results show 'text' and 'item' factors as key considerations. As can also be seen, the 'Other' category was a frequently coded category though no insight was provided as the nature of the responses (Figure 2).
Based on the results of Dinsmore and Parkinson's (2013) investigation, they concluded that:

> It is clear from these data that participants were taking into account multiple factors when rating their confidence. Not only do these data support using Bandura's (1986) model of reciprocal determinism, but further, indicate that inferences made about participants' calibration must be carefully considered. For instance, there were many individuals who only used one factor for making their confidence judgments, while others used different factors for different items. Further some individuals were able to incorporate multiple factors when making their confidence judgments. Thus, one point that is abundantly clear from this study is that judgments of confidence are based on individual differences and require focused, domain-specific study.

Number of responses for each confidence code category by passage.

| Category | SA passage | SIpassage |
|---|---|---|
| Prior Knowledge | 12 | 11 |
| Text characteristic | 49 | 29 |
| Item characteristic | 24 | 25 |
| Guessing | 2 | 8 |
| Other | 25 | 31 |
| Prior knowledge and text characteristics | 5 | 14 |
| Prior knowledge and item characteristics | 9 | 6 |
| Prior knowledge and guessing | 0 | 1 |
| Text characteristics and item characteristics | 6 | 7 |
| Text characteristics and guessing | 1 | 2 |
| Item characteristics and guessing | 4 | 2 |
| Item characteristics and other | 1 | 1 |
| Prior knowledge, text, and item characteristics | 4 | 2 |
| Text and item characteristics and guessing | 0 | 1 |
| Prior knowledge, item characteristics, and guessing | 0 | 1 |

Figure 2. *Bases for confidence judgments for research measures ability from Dinsmore and Parkinson, 2013, p. 11.*

There is some good news in the investigation of this phenomenon: some general tendencies (i.e. not domain specific) have been observed in numerous studies. The prevailing tendencies relate to bias, the directional measure of calibration, which inform us that high proficiency individuals tend to under-estimate their abilities, and low proficiency individuals tend to over-estimate their abilities though hard facts about the underlying nature of misestimations remains to be determined. Regrettably, instructional interventions have not shown to be particularly effective in increasing calibration accuracy. Ultimately, this takes us to a core issue regarding the value of investigating any construct: Why is calibration and the understanding of the nature of calibration important? Calibration is an indirect measurement of deeper thinking. Deeper thinking is a by-product of reflection and metacognition. Calibration accuracy is a by-product of all of the above. It informs the field of learning and instruction opportunities of the holy grail of many educators as it speaks of learner autonomy. Well-calibrated learners apply different strategies in order to solve different tasks, they actively think about what they are doing and why they are doing it (Zimmerman, 2009). Thus, they will make decisions of continuing to work on a task, or seeking new information, or seeking expert assistance in order to move forward. These learner decisions represent an internal understanding of how one learns best. As an introduction to calibration accuracy of oral presentation skills, this paper posed the following research questions:

RQ 1: Are low-level participants well calibrated in oral presentation tasks as measured by self and subsequent instructor assessment?

RQ 2: Will general tendencies listed in extant literature regarding learner bias be supported working with low-level participants?

**Methodology**

**Participants:**

The participants of this study were fifteen 1st-year, female students enrolled in a required English communication class with speaking and writing as its focus. This class is instructed entirely by native speakers and meets for ninety minutes once a week. However, for the oral presentation component, only 10 of the enrolled fifteen participants attended. The class was streamed though all of the participants were classified as low-beginners as per their English placement test. As with all classes, motivation levels were mixed as was reflected in the spotty attendance of several participants particularly during performance events.

**Instrumentation:**

A 5-point scoring scale was used for each of the 4 factors: eye-contact, fluency, gestures and presentation length by which participant oral presentations were assessed. However, confidence scores based on performance outcomes ranged from 0 – 100 to reflect scaling used in other research studies (i.e. precision considerations) with performance standards explicated for every score (See Appendix).

**Procedure:**

The oral presentations on which this study is based occurred in the fall semester, however, the study was rooted in speaking performances that occurred the semester prior. In the spring semester, as part of the participants' course obligations, they performed 2 group discussions. The discussion themes were topics from the textbook, personal in nature, and thus suitable for these low-level learners. The group discussion had the following format: a. brief personal introduction b. 3 main body points were

to be presented lasting at least one minute c. an abbreviated conclusion, and d. Q & A session of 1 minute. The discussion was to be 4 – 5 minutes in length, performed 3 times, and each time to a different group/audience. Learners were scored based on a format similar to the instrument applied in the fall semester. The only difference was that gestures were not included in the group discussion as the learners were allowed to remain seated. Learners were given copies of their scores sheets, which also contained very brief comments. These group discussions acted as the necessary scaffold for participants to make a priori judgments of a given task: the oral presentation.

The fall semester is a continuation of the spring semester and is designed to increase learner speaking output and included 3 presentations. The data for this study was collected for the 1st presentation only. The fall semester opens with the course outline and dates of classes and key events such as quizzes, tests and oral presentations. The learners in the class, at this point, become participants of this study. Oral presentation practice also begins in the first class, and a script for the introduction, main body transitions and conclusion is provided. As a group, modelled by the instructor/researcher, the standard parts of the presentation are rehearsed including the setting of expectations for each of the factors to be investigated. Then, in pairs, participants practice with one person holding the script, and the other, hands free, practices. This process occurred 7 times prior to the oral presentation performance. Further, a presentation planning sheet was provided and classroom activities acted as scaffold to each main body presentation point. These points of reflection are loosely modeled after the factors of consideration, as listed by Dinsmore and Parkinson, 2013, as noted in Figure 2.

As part of the participants' pre-test, they were provided with a copy of their group discussion score sheets for review. Then, they were asked to think about what their performance in group discussion including how much preparation they did, how they felt (i.e. confidence), and what they did for preparation (i.e. work with a partner, practice in front of the mirror, record their speaking event). Finally, the participants were asked based on their experience, practice and watching the teacher modeling what they believed their score would be with one-hundred percent certainty, Participants' a priori judgments were recorded on the same Grade Summary Sheets used by the instructor. As with the process in the spring semester, participants made presentations 3 times, though due to time constraints, there was no Q & A session to

follow each speaking event.

The post-test was the instructor evaluation based on the scoring rubric presented in the Grade Summary Sheet. Recall that instructor modeling and expectations of performance associated with scoring was explicated to the participants repeatedly. Instructor scoring progressed incrementally with every presentation with the final presentation (3rd presentation) being the ultimate score determinant as first presentations often demonstrated more anxiety than the final. Participants received a copy of their presentation scoring sheet to compare their a priori judgments to the instructor's, which also contained several brief comments.

## Results

The participant a priori judgments of performance (pre-test) mean scores are presented along with the demonstrated ability judgments of the instructor (post-test) as well as the associated standard deviations (SD). Participant bias of 3 factors: eye-contact, gestures, and duration were over-estimated abilities based on participant pre-test judgments, while fluency scores were under-estimated (Table 1).

Table 1 *Pre-test and post-test descriptive statistics.*

| Scores | Pre-test | | | Pre-test | |
|---|---|---|---|---|---|
| | M | SD | | M | SD |
| Eye contact | 83.00 | 6.75 | | 67.00 | 17.03 |
| Gestures | 84.00 | 6.99 | | 64.00 | 13.50 |
| Duration | 84.00 | 5.27 | | 64.00 | 14.94 |
| Fluency | 74.00 | 6.99 | | 79.00 | 9.94 |
| Average Learner Perceived Ability | 81.50 | 21.71 | Average Instructor Ability Evaluation | 70.75 | 44.98 |

*Note. N=10.*

Table 2 *Paired samples t-tests for significant differences between pre- and post-test scores.*

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|
| | Mean Difference | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | Lower | Upper | | | |
| Eye contact 1 - Eye contact 2 | 16.00 | 17.13 | 5.42 | 3.75 | 28.25 | 2.95 | 9 | .016 |
| Fluency Skill 1 - Fluency skill 2 | -5.00 | 10.80 | 3.42 | -12.73 | 2.73 | -1.46 | 9 | .177 |
| Gestures 1 - Gestures T2 | 20.00 | 15.64 | 4.94 | 8.82 | 31.18 | 4.05 | 9 | .003 |
| Time 1-Time 2 | 12.00 | 14.76 | 4.67 | 1.44 | 22.56 | 2.57 | 9 | .030 |
| Total 1–Total 2 | 43.00 | 43.47 | 13.75 | 11.90 | 74.10 | 3.13 | 9 | .012 |

Participants scores pre-test and post-test were investigated for significantly different means scores (Table 2). Specifically, there were a significant differences in the score for Eye contact ($M$ = 16, $SD$ = 17.13), $t(9)$ = 2.95, $p$ = .05; Gestures ($M$ = 20, $SD$ = 15.64), $t(9)$ = 4.05), and Time (presentation duration) ($M$ = 12, $SD$ = 14.6), $t(9)$ = 2.57). However, Fluency was not found to have a significant difference though it was approaching: ($M$ = -5.00, $SD$ = 10.8), $t(9)$ = 1.46).

**Discussion**

RQ 1: Are low-level participants well calibrated in oral presentation tasks? The data for this sample would suggest for 3 factors: eye-contact, gesture use and presentation duration, “No, they are not.” Even after previous experience with a similar task accompanied by an attribution of score activity as well as extensive practice inside the classroom the participants were generally off from how well they believed they would perform. There was no follow up with the participants as to what reasons they attached to the mis-calibration though most seemed unaware of how well or poorly they actually performed. However, as regards fluency, the participants were reasonably accurate in their a priori estimations. Generally, they understated their ability though it must be known that a considerable weight for fluency assessment was given to scripted areas of the oral presentation: introduction, transitions and conclusion over participant created main body components. What is clear is that the participants needed more

practice on their own either independently, with a classmate or group of classmates. Perhaps, presentation teams would be an idea to raise out-of-class effort. Peer-modeling in collectivist cultures may have a large impact on learner behavior as they do not want to let their group down. Further, as a classroom activity, planning, designing and implementing regular activities of reflection may increase learner attributions for performance as well as an understand of capabilities of performing other tasks. Reflection is one means for learners to start thinking about thinking (i.e. metacognition).

RQ 2: Will general tendencies listed in extant literature regarding learner bias be supported working with low-level participants? The data from this study suggests, "Yes", the participants demonstrated a bias that is consistent, generally, with the extant literature on the phenomenon. The literature on calibration is clear in that individuals with low ability tend to overstate their performance capabilities as was evidenced in this study. All of the participants are classified as low-level English language learners; hence, it is hypothesized that they would have a positive bias, or in other words, they would over-estimate their performances, which is what occurred. Once again, the reasons for this are unclear though the literature does indicate general cognitive ability, achievement and general academic ability as possible determinants. These determinants further imply that development of metacognitive practices with reflection being a key component may raise one's ability at understanding what their ability actually is and how they will perform based on what the task is, knowing the standards of evaluation, and knowing their performance will also be a function of what they do to prepare, and how they do it.

## Conclusion

Calibration is a difficult concept to measure. However, with growing interest, continued research and advanced statistical methods, there is a possibility that more light will shine on and expose some of the many mysteries of individual differences, which are helpful in predicting performance and ultimate achievement of a range of learners in a range of domains. Currently, there are more questions than answers in this area of study but leading learners to think about their abilities and subsequent performance in academic tasks is an academically profitable exercise because the more accurate learners are in estimating their perceived ability to actual ability demonstrated is an indirect measure of how much metacognitive practice a learner has engaged. The benefits of the latter are well established in learning and instruction, yet calibration remains a work in progress.

**Bibliography**

Alexander, P. A. (2013). Calibration: What is it and why it matters? An introduction to the special issue on calibrating calibration. *Learning and Instruction,* 24, 1-3.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review,* 84(2), 191-215

Bandura, A. (1986). *Social foundations of thought and actions: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of child development. Six theories of child development (pp. 1-60).* Greenwich, CT: JAI Press.

Bandura, A. & Adams, N. E. (1977). Analysis of self-efficacy theory of behavioral change. *Cognitive Therapy and Research, 1*(4), 287-310

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5,* 307-337.

Bandura, A. et. al. (1996). Multifaceted Impact of Self-Efficacy Beliefs on Academic Functioning. *Child Development, 67,* 1206-1222.

Bandura, A. (1997). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in Changing Societies (pp. 1-45).* Cambridge: Cambridge University Press.

Boekaerts, M. and Rozendaal, J.S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students' accuracy of feeling of confidence. *Learning and Instruction,* 20, 372-382.

Bol, L., & Hacker, D.J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education, 69 (2),* 133-152.

Brannick, M.T., Miles, D.E. & Kisamore, J.L. (2005). Calibration between Student Mastery and Self-efficacy. *Studies in Higher Education,* 30(4), 473-483.

De Grez, L., Valcke, M. and Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education,* 13, 129.

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.

Hadwin, A.F., & Webster, E.A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and Instruction,* 24, 37-47.

Flavell, J.H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist,* 34(10), 906-911.

Green, S & Salkind, N. (2008). *Using SPSS for Windows and Macintosh: Analyzing and understanding data (5th ed.)*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kruger, J. & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology,* 77(6), 1121-1134.

Lundeberg, M. A., Fox, P. W., & Punćcohaŕ, J. (1994). Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of Educational Psychology, 86*(1), 114.

Oettingen, G. (1995). Cross-cultural perspective on self-efficacy. In A. Bandura (Ed.), Self-efficacy in changing societies (pp. 149-176). New York: Cambridge University Press.

Pajares, F. (2002). Overview of social cognitive theory and of self-efficacy. Retrieved December 26, 2015, from http://www.emory.edu/EDUCATION/mfp/eff.html

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology, 82(1)*, 33-40.

Schunk, D. & Parajes, F. (2009). In K. Wentzel & A. Wigfield (eds). *The Handbook of Motivation in School 1st ed. (pp. 35-54)*. New York: Routledge.

Tabachnick, B. & Fidell, L. (2007). *Using Multivariate Statistics (5th ed.)*. Boston: Pearson Education.

Wang, C., Kim D., Bai, R. and Hu, J. (2014). Psychometric properties of a self-efficacy scale for English language learners in China. *System,* 44, 24-33.

Zimmerman, B. (1995). Self-efficacy in educational development. In A. Bandura (ed.). *Self-efficacy in Changing Societies (pp. 202-231)*. Cambridge: Cambridge University Press.

Zimmerman, W.A. (1997). Self-efficacy in educational development. In A. Bandura (Ed.), *Self-efficacy in Changing Societies (pp. 202-231)*. Cambridge: Cambridge University Press.

Zimmerman, B. (2000). Self-Efficacy: An Essential Motive to Learn. *Contemporary Educational Psychology 25,* 82-91.

## Appendix

## Presentation Scoring Rubric

| Score | Eye-contact | Score | Natural English/Voice | Score | Gestures | Score | Preparation/Time |
|---|---|---|---|---|---|---|---|
| 5 (90-100) | Maintained for 75% + of the presentation. | 5 (90-100) | Speaks confidently, clearly and smoothly for 75%+ of the presentation using a variety of fluency skills (linking, contractions, intonation) throughout the presentation. | 5 (90-100) | Presenter uses gestures throughout presentation including the use of hands and finger to indicate progression and sequence. | 5 (90-100) | Presenter has shown they have done proper research and practiced their presentation. Prescribed format followed perfectly. 4 hours + of preparation evidenced. Presentation remains unchanged from first presentation to last. 5 minutes in length. |
| 4 (80-89) | Maintained for approximately 60%+ of the presentation. | 4 (80-89) | Speaks confidently, clearly and smoothly for the introduction and conclusion and parts of the body of the presentation using a variety of fluency skill (linking, contractions, intonation). | 4 (80-89) | Presenter uses gestures to indicate progression and sequencing, and sometimes uses gestures in other parts of the presentation. | 4 (80-89) | Presenter has done proper research but exhibits nervousness and makes some errors (stops presentation to find speaking point), which means that more presentation practice was needed. Prescribed format followed perfectly. 3+ hours or preparation evidenced. 4 minutes in length. |
| 3 (70-79) | Maintained approximately 40% of the presentation. | 3 (70-79) | peaks confidently, clearly and smoothly for approximately 40% of the presentation using a variety of fluency skill (linking, contractions, intonation) | 3 (70-79) | Presenter sometimes uses gestures to indicate progression and sequence. | 3 (70-79) | Prescribed format almost perfectly followed and nervousness exhibited. Some direct cut and paste from internet. 2+ hours of preparation evidenced. 3 minutes in length. |
| 2 (60-69) | Reads presentation 75%+ of the presentation. | 2 (60-69) | Speaks many words separately, slowly, has many pauses, and speaks without intonation. | 2 (60-69) | Presenter occasionally uses gestures to indicate progression and sequence. | 2 (60-69) | Prescribed format generally followed, but some points not covered. Many pauses during presentation. Many parts copies from internet. 1+ hours of preparation evidenced. 2 minutes in length. |
| 1 (50-59) | Reads presentation more than 90% of the time. | 1 (50-59) | Speaks with many pauses, little intonation or links, contractions, etc. | 1 (50-59) | Presenter rarely uses gestures to indicate progression and sequence. | 1 (50-59) | Prescribed format not followed: many points not covered. Many pauses during presentation. Many parts copied from internet. Insufficient practice evidenced. 1 minute in length. |
| 0 No score | Reads presentation. | 0 No score | Not able to hear or understand many parts of the presentation. | 0 No score | No gestures used. | 0 No score | Prescribed format generally followed, but some points not covered. Many pauses during presentation. Copied from internet. No preparation evidenced. Less than 1 minute in length. |