

A Rasch analysis of L2-English achievement goals of female university students

Jean-Pierre Joseph Richard

Introduction

Achievement goals are cognitive representations of future-focused purposes learners adopt in academic situations to direct behavior to approach or avoid competence-related end states (Hulleman, Shrager, Bodmann, & Harackiewicz, 2010). The current study employed the 2×2 achievement goals model, with two achievement goals (*i.e.*, mastery, performance) and two valence dimensions (*i.e.*, positive, negative). This model resulted in the Achievement Goals Questionnaire (AGQ) and revised form (AGQ-R) (Elliot & McGregor, 2001; Elliot & Murayama, 2008), with items intended to measure four achievement goal orientations: (1) mastery approach (MAp); (2) performance approach (PAp); (3) mastery avoidance (MAv); and (4) performance avoidance (PAv). Mastery goals relate to competence development and skill acquisition, while performance goals have both positive and negative consequences, and refer to the displaying of knowledge or ability. Positive valence refers to approaching success; negative valence refers to avoiding failure.

One criticism of achievement goals is that researchers have primarily depended upon confirmatory factor analysis (CFA). Murayama, Elliot and Yamagata (2011) responded to this criticism by devising five unique studies employing different procedures and methodologies; however, they used these

with the PAp and PAv subscales exclusively. There have been a negligible number of achievement goals studies using the Rasch Model (RM). Muis, Winne and Edwards (2009) were the first to research achievement goals using both CFA and RM principal components analysis (PCA). They found the CFA replicated previous research results; however, the RM indicated poor reliability for individuals. Muis, Winne and Edwards claimed that the AGQ might well measure the goal orientations of the total sample but not necessarily of the individual. Moreover, they argued that one reason the data poorly fit the model is the limited number of items ($k = 3$) measuring each distinct goal, and thus proposed additional items be added to the questionnaire. Importantly, although Muis, Winne and Edwards tested the RM on the four hypothesized goals, MAp, PAp, MAV and PAV; they did not report using RM on competing models composed of possible combinations of mastery, performance, approach or avoidance goals.

Hart, Mueller, Royal and Jones (2013) investigated the AGQ-R with two distinct populations of African American high school students, rural and urban, using both CFA and RM. In the rural sample, the 2×2 model had the best fit but it did not have good fit according to fit indicators; and in the urban sample, no model had good fit using CFAs. Hart et al combined the complete set of 12 items from the AGQ-R, which are theoretically intended to measure four unique goals, when using the RM. They erroneously identified one unidimensional construct for each sample; erroneous because the first extract accounted for significantly less than half of the variance in both populations which should have indicated the presence of further dimensions. Similar to Muis, Winne and Edwards (2009), item reliabilities were generally high but person reliabilities were lower. The lower person reliabilities might be a reflection of the limited number of items measuring goals on each subscale.

Research Question

Using RM to investigate the 2×2 model of achievement goals, different researchers have reached different conclusions. Muis, Winne and Edwards (2009) identified four distinct goal orientations, and thus argued that the 2×2 model is multidimensional. In contrast, Hart et al (2013) claimed the 2×2 model is unidimensional. Moreover, low person reliabilities have been found, which might be a result of too few items measuring each goal. In this current study, the AGQ-R was expanded by adding two additional items per goal to create a 20-item survey. The primary purpose of this study is to begin the validation of this expanded questionnaire using RM principal component analysis.

Methodology

Participants

Data were collected for a large-scale longitudinal mixed-methodology research project over one academic year, beginning in April, 2013 from participants at 13 national and private universities in Japan. Participants described in this current paper ($N = 125$) were extracted from the larger study, and are from several departments at one all-female university in the Kantō region. Department standardized T-scores (偏差値) ranged from high-40s to mid-50s. Before the beginning of this study, the participants had been studying English for a minimum of six years. At the time of the study, the participants were enrolled in six sections of a first-year English communication class taught by three native-English speaker lecturers. All participants were explained the purposes of the study and consented to participate.

Procedure

In April, 2013, participants completed several documents including a

consent form and an expanded version of the AGQ-R. The AGQ-R ($k = 12$) has three items per goal, the stems of which are *My aim is...*, *I am striving...*, and *My goal is...*. In this current paper, the AGQ-R was expanded with two additional items per goal, with the following stems: *I work toward...*, and *My target for...*. Additional sample items included: *I work toward becoming competent in this class* (MAp); *My target is for my performance to be better than others in this class* (PAp); *My target is to avoid not learning in this class* (MAv); and *I work toward avoid being worse than others in this class* (PAv). Each of the 20 items referred to *this English class* and were scored on a six-point Likert scale: 1 = strongly disagree; 6 = completely agree, with no neutral mid-point.

Analyses

RM, which was used to test the validity of the expanded AGQ-R in this study, estimates the probability that a participant selects a certain response category for a particular item. Linacre (2002) suggested several criteria for evaluating rating scale effectiveness. First, there should be at least ten observations per scale category. Second, average measures should advance with each successive category, that is the second category (*e.g.* disagree) should be more difficult than the first category (*e.g.* strongly disagree). Third, outfit mean statistics, sensitive to outliers, should be <2.0 . Fourth, Andrich thresholds should be ordered and advance by more than 1.4 logits (logarithm of odds unit) but less than 5.0 logits per category. Following Andrich (2013) categories were collapsed to avoid disordered thresholds. A threshold is disordered when a higher-ranked item category (*i.e.*, theoretically more difficult) is more easily endorsed than a lower-ranked item category; for example, if the item difficulty of a category four (somewhat agree) is lower (*i.e.*, easier) than the item difficulty of a category three (somewhat disagree).

Complex data matrices are reduced to one unidimensional variable in RM:

all data are explained by one latent variable, and the remaining unexplained data is presumed to be random noise. Linacre (2007, 2012) has suggested four criteria for evaluating unidimensionality: (1) the explained variance is greater than 50%; (2) the first remaining contrast should be less than 3.0 eigenvalues (EV), and or accounting for less than 10% of the unexplained variance; (3) item loadings, either positive or negative, greater than .40 are considered to be substantive, and disattenuated correlations between contrast clusters of item loadings should be greater than or equal to .82; and (4) the factor loadings should be investigated for meaningfulness. If the positive and negative factor loadings appear to be partitioned separately into meaningful structures, then these structures merit further investigation. Moreover, Hagell (2014) stressed that interpretation of the PCA results should be based on variable definitions (*i.e.*, construct theory).

In RM, the item-person map (Wright variable map), which places person and item measures on a common scale showing hierarchy and location relative to one another, should be visually inspected. Persons and items are placed vertically, with the former on the left of the map and the latter on the right; the top of the map indicates more person ability (or more item difficulty) and the bottom of the map indicates less person ability (or less item difficulty). Along the scale are the letters M, S and T which correspond to mean, one standard deviation and two standard deviations; and person and item means should be close (*i.e.*, less than two measurement errors apart). Greater separation between persons and items likely indicates that the sample of participants are not well matched to the instrument. In this paper, the Wright variable map has been replaced by the Rasch-Thurstone threshold map which replaces the items with the range of coverage of the item categories.

Results

Unadjusted means (standard deviations) for the four achievement goals for this sample ($N = 125$) were highest for MAp, 4.43 (0.98); followed by PAv, 4.13 (0.93); MAV, 4.05 (1.02); and PAp, 3.90 (1.07) respectively. Mean scores for MAp were statistically significantly higher than the remaining three goal orientations; the mean scores for the remaining three goals were not statistically significantly different from each other.

All 20 items were initially tested as Model A for unidimensionality; person (.93) and item (.94) reliabilities were strong. Certain rating scale categories needed to be combined (*i.e.*, collapsed) to ensure there were at least ten observations per scale category, and two items, MAp2 and MAp5 had to be deleted because of large numbers of unexpected responses (>10%). Of the remaining 18 items, average measures advanced successively, outfit means square statistics were <2.0, and Andrich thresholds advanced appropriately. Total variance explained was 51.6%, which is above the 50% minimum; however, the first contrast EV was large (3.3), the disattenuated correlation between contrasts 1 and 3 was below the cutoff. Finally, the set of loadings were scrutinized. The positive loadings >.40 included four PAp items and one PAv item; and the negative loadings included four MAV items. Taken together, the full questionnaire ($k=20$, reduced to $k=18$) does not appear to be unidimensional.

Second, five additional models were also tested for unidimensionality: Model B: All-Mastery, MAp and MAV goals; Model C: All-Performance, PAp and PAv goals; Model D: All-Approach, MAp and PAp goals; Model E: All-Avoidance, MAV and PAv goals; and Model F, the Trichotomous Framework, MAp, PAp and PAv goals. Due to the large first contrast EV, the percentages of variance in the first contrast, the disattenuated correlations below the cutoff for unidimensionality, and the generally clear theoretical distinction

between the items which loaded positively and negatively, all five of these models were also found to be multidimensional. Results for Models A through F are summarized in Table 1.

Third, individually, the four achievement goals, each with five-items, were tested for unidimensionality. Results are summarized in Table 2. Person and item reliabilities were moderate to strong. Infit and outfit mean squares were within good ranges. Andrich thresholds rose appropriately for three of the goals; however, for PAp goals, two of the thresholds were large. For three of the goals, the percentage of raw variance explained was greater than .50; EV in first contrasts were less than 3.0. For PAp goals, and possibly MAp goals, the percentage of first contrast variance was acceptable. Between three and five items loaded greater than .40 and disattenuated correlations were near good to high except for PAv goals. The following sections describe results of each achievement goal separately.

Mastery Approach (MAp)

Categories were combined to ensure a minimum of ten observations per category; average measures were inspected for advancement with each successive category; outfit mean statistics were viewed to verify they were less than 2.0; and Andrich thresholds were scrutinized that they advanced by more than 1.4 logits. This was repeated for all goals. The MAp goals Rasch-Thurstone threshold map is displayed in Figure 1. MAp2, *My goal is to learn as much as possible in this course*, category 3 (somewhat disagree) is easiest to endorse; and *My aim is to completely master the material presented in this class*, category 6 (strongly agree) is most difficult. Surveying Figure 1, it can be seen that the mean difficulty of the items, the M at 0 logits, is below the mean ability of the respondents, the M at 1.2 logits, meaning it is easier for the participants to agree to the items. Overall, there is a good spread of participants and item categories; however, no students are targeted by the

Table 1. Summary of Rasch PCAs for Models A through F.

	Model A All-Items	Model B All-Mastery	Model C All- Performance	Model D All-Approach	Model E All- Avoidance	Model F Trichotomous
Person reliability	.93	.87	.92	.88	.83	.92
Item reliability	.91	.96	.88	.97	.76	.96
Infit MNSQ	.67 to 1.39	.59 to 1.51	.67 to 1.47	.58 to 1.36	.69 to 1.32	.70 to 1.44
Outfit MNSQ	.65 to 1.44	.59 to 1.48	.62 to 1.43	.55 to 1.51	.71 to 1.26	.70 to 1.38
Andrich thresholds	-4.08	-3.33	-5.68	-4.27	-2.36	-4.03
per category ¹	-1.50	-1.65	-1.92	-1.00	-.16	-1.18
somewhat agree	0.95	1.11	1.14	1.47	2.52	1.24
strongly agree	4.63	3.87	6.46	3.80	6.46	3.97
Raw variance explained	51.6	54.9	61.6	59.9	42.3	55.6
First contrast	3.3	2.3	2.8	2.6	2.8	3.3
Eigenvalue EV	8.9	10.2	10.7	10.5	16.0	9.6
% 1st contrast variance						
Positive loadings	PAP4 .68	MAV5 .66	PAV3 .76	PAP2 .79	MAV4 .69	PAP2 .76
≥.40	PAP2 .60	MAV4 .64	PAV1 .75	PAP3 .66	MAV5 .67	PAP3 .69
	PAP3 .55	MAV1 .43	PAV4 .69	PAP5 .65	MAV1 .45	PAP4 .69
	PAP1 .48			PAP4 .43	MAV2 .42	PAP5 .52
	PAV5 .48					PAP1 .44
Negative loadings	MAV5 -.59	MAp3 -.61	PAP4 -.59	MAp4 -.55	PAV5 -.59	PAV5 -.55
≥.40	MAV4 -.57	MAp4 -.57	PAP1 -.57	MAp3 -.54	PAV3 -.58	PAV4 -.51
	MAV1 -.49	MAV3 -.57	PAP2 -.55	MAp5 -.51	PAV4 -.52	PAV3 -.42
	MAV2 -.47		PAP3 -.40		PAV1 -.50	
Disattenuated correlations	.76	.76	.72	.80	.58	.77

Note: 1. For Models A through F, categories 1 (strongly disagree), 2 (disagree) and 3 were collapsed into category 3 (somewhat disagree) for all items due to the limited number of respondents in each of categories 1 and 2. For Model E, categories 5 (agree) and 6 (strongly agree) were also collapsed into category 5.

Table 2. Summary of Rasch PCAs for the Four Achievement Goals.

	Mastery Approach (MAp)	Performance Approach (PAp)	Mastery Avoidance (MAv)	Performance Avoidance (PAv)
Person reliability	.75	.91	.77	.76
Item reliability	.96	.90	.79	.76
Infit MNSQ	.80 to 1.20	.76 to 1.40	.62 to 1.20	.70 to 1.36
Outfit MNSQ	.78 to 1.20	.69 to 1.56	.62 to 1.33	.68 to 1.38
Andrich thresholds per category ¹	-3.52 -1.46 1.52 3.46	-8.26 -1.86 2.79 7.33	-2.44 -.47 2.92	-1.87 1.87
Raw variance explained	56.5	78.1	52.3	43.6
First contrast Eigenvalue	1.5	1.8	2.1	2.2
% of 1st contrast variance	12.9	7.9	19.6	25.2
Positive loadings	MAp1 .87	PAP1 .80 PAP4 .60	MAv4 .77 MAv5 .77	PAv2 .82 PAv5 .81
≥.40				
Negative loadings	MAp5 -.62 MAp4 -.52	PAP3 -.58 PAP5 -.50 PAP2 -.44	MAv2 -.69 MAv1 -.63	PAv1 -.69 PAv3 -.52 PAv4 -.41
≥.40				
Disattenuated correlations	.76	.92	.76	.46

Note: 1. For all four goal categories, categories 1 (strongly disagree) and 2 (disagree) were collapsed into category 2 due to the limited number of respondents in category 1; and this was repeated for categories 2 and 3 for PAV goals. Also, for MAv and PAV goals, categories 5 (agree) and 6 (strongly agree) were collapsed into category 5.

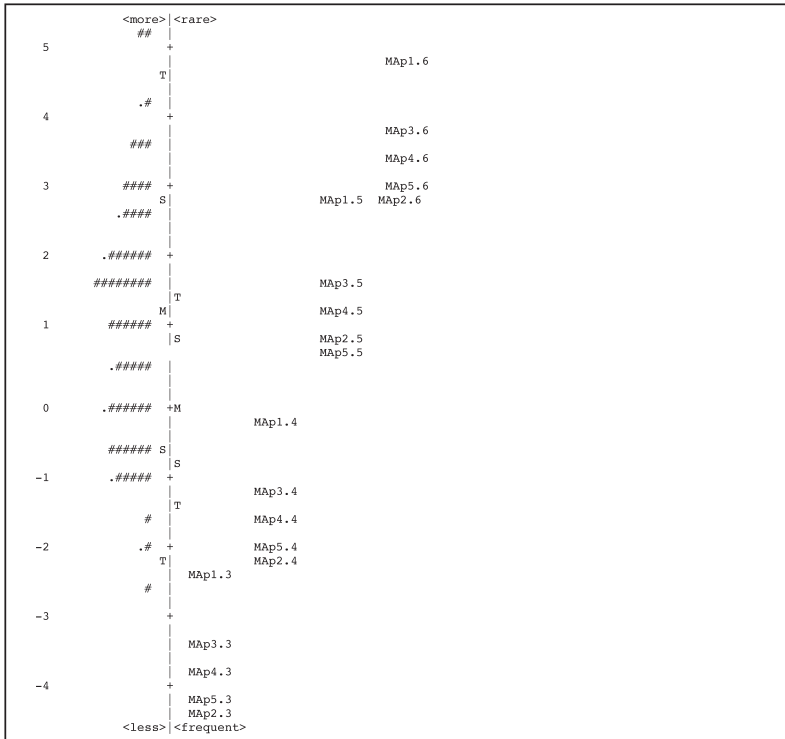


Figure 1. Mastery Approach Measure Person Map Item 50% Cumulative probabilities (Rasch-Thurstone thresholds). Each “#” is 2 participants, each “.” is 1.

easiest item categories.

Performance Approach (PAp)

Figure 2 displays the PAp goals Rasch-Thurstone threshold map. PAp3, *My goal is to perform better than the other students*, category 3 (somewhat disagree) is easiest; PAp1, *I am striving to do well compared to other students*, category 6 (strongly agree) is most difficult. It can be seen that mean difficulty of items and persons ability are equal, the M+M at 0 logits. There is a good spread of participants and item categories; however, no students are

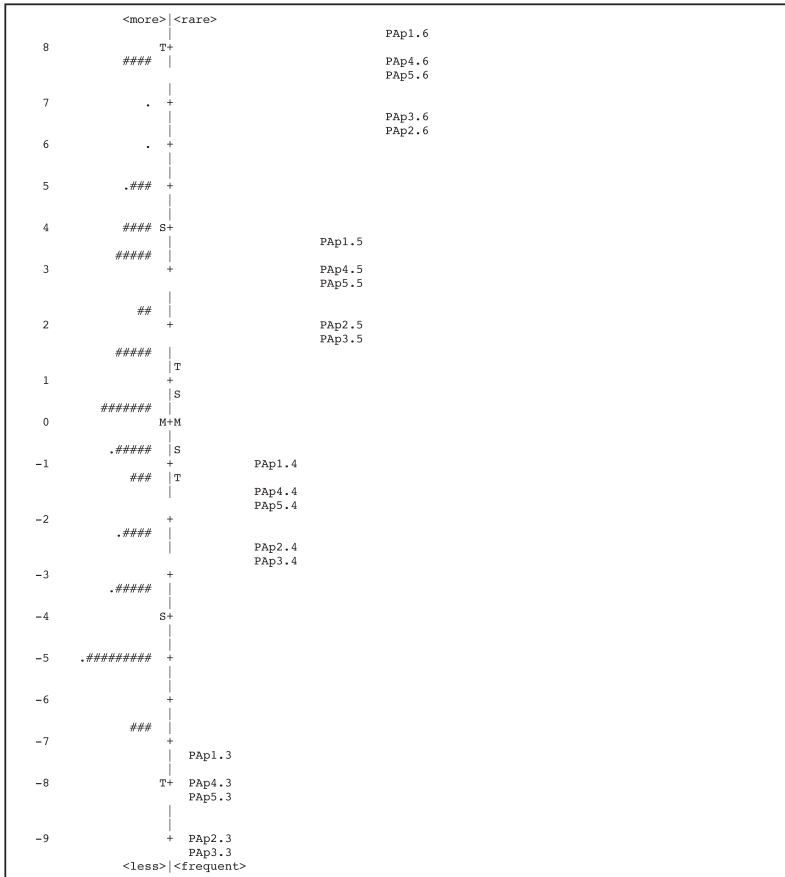


Figure 2. Performance Approach Measure Person Map Item 50% Cumulative probabilities (Rasch-Thurstone thresholds). Each “#” is 2 participants, each “.” is 1.

targeted by the easiest item categories and few targeted by the most difficult ones. Lower items are too easy for this group and higher ones are too difficult. More average-difficulty items are needed.

Mastery Avoidance (MAv)

The MAv goals Rasch-Thurstone threshold map is displayed in Figure 3.

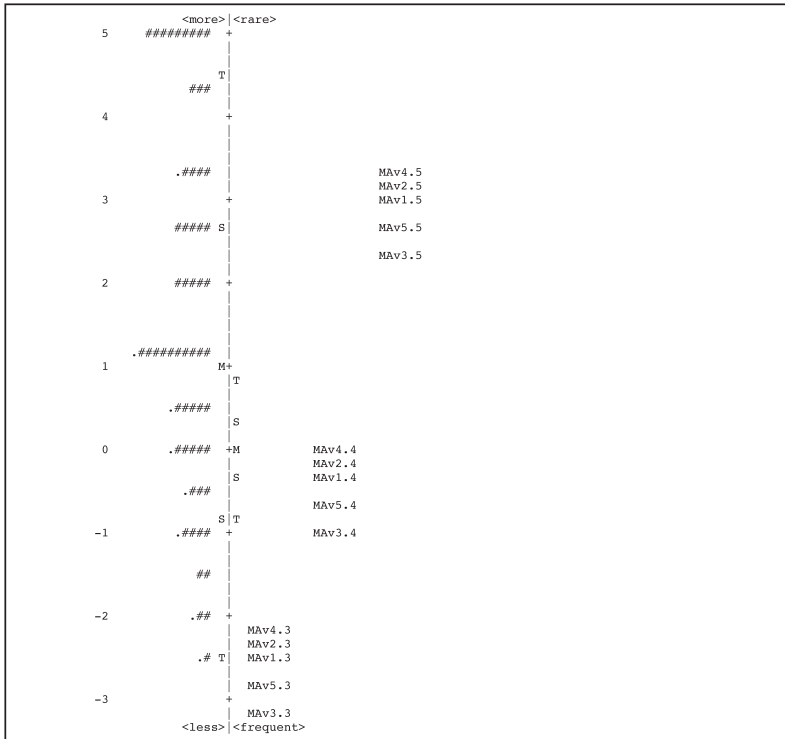


Figure 3. Mastery Avoidance Measure Person Map Item 50% Cumulative probabilities (Rasch-Thurstone thresholds). Each “#” is 2 participants, each “.” is 1.

MAv3, *I am striving to avoid an incomplete understanding of the course material*, category 3 (somewhat disagree) is easiest to endorse; MAv4, *I work toward avoiding a misunderstanding of the material in this course*, category 5 (agree) is most difficult. The mean for item difficulties and for person abilities are separated by one logit. Item categories were easy for this group to endorse. There are no item categories for approximately 20% of the participants in this sample between logits 3.5 and 5.0. Simultaneously, there are no categories targeting one-third of the sample between logits 0 and 2.

MAv items need further investigation.

Performance Avoidance (PAv)

The PAv goals Rasch-Thurstone threshold map is displayed in Figure 4. PAv4, *I work toward to avoid being worse than others in this class*, category 4 (somewhat agree) is easiest to endorse; PAv5, *My target is to avoid having a poor performance in this class*, category 5 (agree) is most difficult. Means for item difficulties and person abilities are equal. Critically, while all five items are found on the map, there are few categories. There are no items for the participants with least nor highest ability. Likewise, no items target the majority of the participants, those found between -1.0 logits and 1.0 logits. PAv goals had the weakest amount of raw variance explained, highest first contrast EV and percentage of unexplained variance, and weakest disattenuated correlation. PAv goals need further consideration.

Discussion

In this study, I examined the psychometric properties of a newly extended version ($k = 20$) of the achievement goals questionnaire-revised form (AGQ-R) (Elliot & Murayama, 2008) using the Rasch Model. To extend the questionnaire I added two additional items per goal. I began by testing Model A, composed of all 20 items, later reduced to 18 due to poorer fit of two items, MAp2 and MAp5, for unidimensionality. Subsequently, I tested five other competing models, All-Mastery, All-Performance, All-Approach, All-Avoidance, and the Trichotomous framework, composed of either 10 or 15 items. All six models, A through F, were found to be multidimensional. Lastly, all four achievement goals, mastery-approach (MAp), performance-approach (PAp), mastery-avoidance (MAv), and performance-avoidance (PAv), were tested separately for dimensionality.

All four goals generally fit the model, although certain weaknesses were identified. Whereas person and item reliabilities for PAp goals, and item

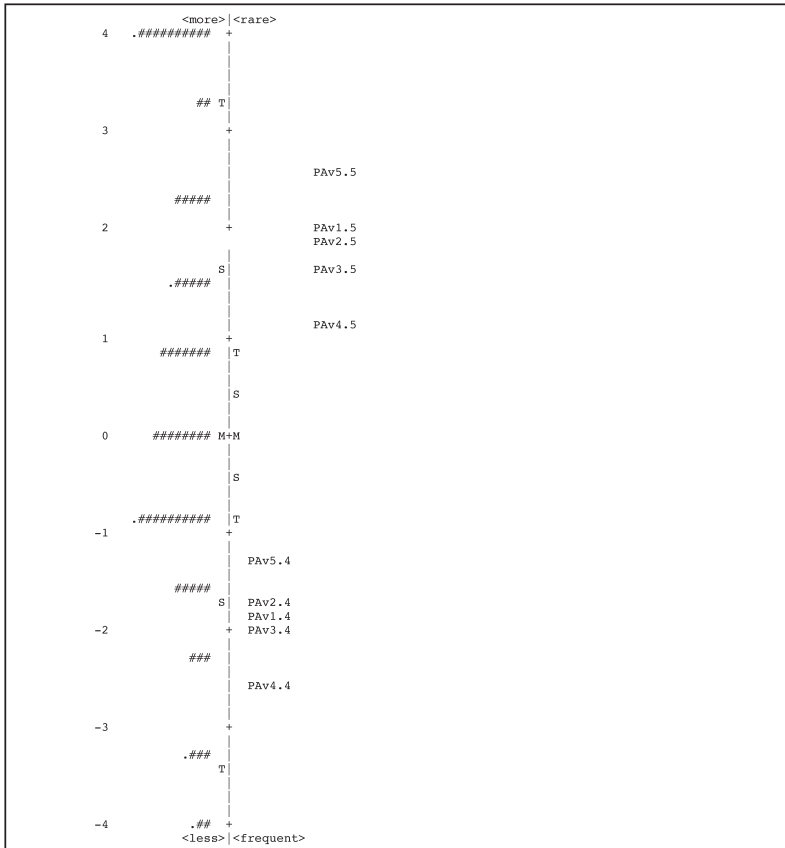


Figure 4. Performance Avoidance Measure Person Map Item 50% Cumulative probabilities (Rasch-Thurstone thresholds). Each “#” is 2 participants, each “.” is 1.

reliabilities for MAP goals were high; person and item reliabilities for the remaining goals might only be considered good. The sample discussed here rarely selected the lowest point of the scales (strongly disagree) for all goals. As a result of this, the lowest categories needed to be combined. Moreover, for PAV goals, the lowest three categories and the top two categories needed to be combined. This indicates that the sample are more likely to endorse the

goals with little variation between them. Consequently, for PAv goals specifically, and all three other goals generally, more items and item categories, with a greater spread from easier to more difficult, would be needed to raise the person reliabilities; and a larger sample of participants with a greater spread of less to more ability would be needed to raise the item reliabilities.

An important issue raised in the literature is the dimensionality of achievement goals. As noted above, Hart et al (2013) and Muis, Winne and Edwards (2009) used different approaches when investigating achievement goals with RM. The former tested dimensionality with all 12 items of the AGQ-R and claimed these items were measuring one unidimensional construct. The problem with the approach by Hart et al is the disconnection between the observed data and the theory underpinning the achievement goal model. Moreover, Hart et al did not test other possible achievement goal models. In contrast, Muis, Winne and Edwards tested each of the four goals separately to identify four different dimensions (MAp, PAp, MAV and PAV); however, they too did not test contrasting models. In this study, I began by testing all 20 items of the questionnaire, and then competing models before I tested each of the four goals separately. In doing so, I was able to examine closely the various potential dimensions of this extended version of the AGQ-R, and consequently, was able to discover that for this sample, the two approach goals, in particular PAp, had the best fit to the model; whereas the two avoidance goals, in particular PAV, had the poorest fit.

Conclusion

The results from the current study provide limited validation evidence for an extended version of the achievement goals questionnaire—revised form which was tested with an all-female sample of first-year students enrolled in a required L2-English communication class in a private university in Japan.

Findings from RM (a) highlight the general strength of the four unique dimensions of the extended version of the AGQ-R; (b) however, the need to write more items with greater difficulty levels is evident.

References

- Andrich, D. (2013). An expanded derivation of the threshold structure of the polytomous Rasch Model that dispels any “threshold disorder controversy”. *Educational and Psychological Measurement, 73* (1), 78–124.
- Elliot, A. J., & McGregor, H. A. (2001). A 2×2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501–519.
- Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*, 613–628.
- Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/t-rest protocol with the Rasch Model: The primacy of theory over statistics. *Open Journal of Statistics, 4*, 456–465. doi: 10.4236/ojs.2014.46044.
- Hart, C. O., Mueller, C. E., Royal, K. D., & Jones, M. H. (2013). Achievement goal validation among African American high school students: CFA and Rasch results. *Journal of Psychoeducational Assessment, 31* (3), 284–299.
- Hulleman, C. S., Shrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin, 136*, (3), 422–449.
- Linacre, J. M. (2002). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3* (1), 85–106.
- Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago, IL: MESA Press.
- Linacre, J. M. (2012). *Dimensionality: Contrasts & variances*. Retrieved from <http://www.winsteps.com/winman/index.htm?principalcomponents.htm>.
- Muis, K. R., Winne, P. H., & Edwards, O. V. (2009). Modern psychometrics for assessing achievement goal orientation: a Rasch analysis. *British Journal of Educational Psychology, 79*, 547–576.
- Murayama, K., Elliot, A. J., & Yamagata, S. (2011). Separation of performance-approach and performance-avoidance achievement goals: A broader analysis. *Journal of Educational Psychology, 103* (1), 238–256.

Keywords

achievement goals, Rasch Model, dimensionality