

Rasch Validating an Instrument Measuring Autonomous Reasons and Achievement Goals

Jean-Pierre J. Richard
The University of Nagano

Reference Data:

Richard, J.-P. J. (2025). Rasch Validating an Instrument Measuring Autonomous Reasons and Achievement Goals. In B. Lacy, M. Swanson, & P. Lege (Eds.), *Moving JALT Into the Future: Opportunity, Diversity, and Excellence*. JALT. <https://doi.org/10.37546/JALTPCP2024-07>

Proximal classroom-based goals and the distal reasons for those goals (*i.e.*, the what and why of classroom learning), are predictors or consequences of many learning-related variables (Sommet et al., 2021). When considered together, these constructs—known as goal complexes—may offer greater explanatory power than goals or reasons alone (Sommet & Elliot, 2017). However, few studies have developed valid instruments to measure both components concurrently. This study aimed to improve item quality for items representing autonomous reasons for goals (Ryan & Deci, 2000) and approach goals from the 3×2 achievement goal model (Elliot et al., 2011) toward the eventual construction of a goal complex framework. Data were gathered at a public university in central Japan where students must demonstrate mastery of the New General Service List (NGSL, Browne, 2013). A questionnaire, composed of items targeting autonomous reasons ($k = 5$), mastery-based goals ($k = 10$), and other-based goals ($k = 5$), was administered to Year 1 students (response rate = 40%). Rasch analysis revealed that the sets of items performed well on most, but not all, of the criteria described by Fisher (2007).

教室内における近接的な目標と、それらに対する遠因的な理由(すなわち、学習の「何を」と「なぜ」)は、多くの学習関連変数の予測因子または結果とされている(Sommet et al., 2021)。これらの構成要素を複合的に考えることで、単独の目標や理由よりも高い説明力を持つ可能性があると考えられる(Sommet & Elliot, 2017)。しかし、これらを同時に測定する妥当なツールの開発はほとんど行われていない。本研究は、自律的な目標の理由(Ryan & Deci, 2000)および3×2達成目標モデル(Elliot et al., 2011)におけるアプローチ目標を表す項目の質を改善し、将来的に複合的な目標の枠組みを構築することを目的とした。データはNew General Service List (NGSL, Browne, 2013)の習得を求められる日本の中部地方にある公立大学で収集された。1年生を対象とした質問紙(回収率 = 40%)のRasch分析の結果は、Fisher (2007)によって示された基準の大部分を満たしてはいたが、すべてではないことが明らかになった。

This paper reports one stage of a long-term project at a regional, public university in central Japan, Shōzan University (SU, a pseudonym) to enhance program-wide materials in order to help learners develop their L2-English competencies and promote positive affective learning behaviours. Corresponding with this, one goal for students in the English language programme at SU was to improve their scores on a commonly used standardized test, the Listening and Reading sections of the Test of English for International Communication (TOEIC L&R). This test is used as an important key performance indicator (KPI) of the English language program at SU. A second goal was for students to develop mastery of the New General Service List (NGSL, Browne, 2013). This paper will first briefly introduce background variables related to this study, the reasons for goals and achievement goals, then introduce the setting and the research. The goal of this current paper is to describe the validation of questionnaire items related to autonomous reasons for study and achievement goals. These items will be later used in a study of goal complexes, which combine proximal classroom-based goals and distal reasons to explain the what and why of classroom learning, in a single item.

Reasons for goals, the why of classroom learning as well as the what of classroom learning, are predictors or consequences of many important variables that relate to learning (Sommet et al., 2021). Reasons from Self-Determination Theory (SDT, Ryan & Deci, 2000) can be separated into two principal types of reasons for goals: autonomous and controlled. Individuals with autonomous reasons for goals are assumed to be more self-determined as they are energized by interest (*i.e.*, intrinsically motivated), or because goal-related activities align with the individuals' identity and values (*i.e.*, integrated regulation), or because the activities are personally relevant and important (*i.e.*, identified regulation). In contrast, individuals with controlled reasons for goals are assumed to be less self-determined in that they are externally regulated, such as by pressure from parents or society, or rewards, such as grades or prizes, or they experience introjected regulation, driven by internalized pressures, such as guilt or a sense of obligation.

JALT2024 • MOVING JALT INTO THE FUTURE: OPPORTUNITY, DIVERSITY, AND EXCELLENCE

NOVEMBER 15–18, 2024 • SHIZUOKA GRANSHIP, SHIZUOKA, JAPAN

Achievement goals are the aims of learners' classroom related competence-relevant motivational behaviours (Elliot & McGregor, 2001). Various models of achievement goals have been proposed, including a 3×2 model (Elliot et al., 2011) composed of self-based competence goals, in which an individual compares their current and previous ability. Another is task-based absolute mastery goals, where the focus is on knowing and understanding items on the current tasks. A third would be performance-based goals (or other-based goals), where the focus is on one's performance vis-à-vis their peers' performance; with two directions of goal-related behaviors. Lastly, approach goals (striving for success) and avoidance goals (trying to avoid failure) are the last models.

Richard (2022) reported on the results from a principal components analysis of Rasch residuals for survey items intended to measure variables from the 3×2 model. The results suggested two dimensions, that is mastery and performance goals, or three, that is self-, task-, and other-based goals, without a clear separation of approach and avoidance directions. In a second study, Richard (2025) investigated items intended to represent approach goals only, that is avoidance-based goals were not investigated, from the 3×2 model of achievement goals, autonomous and controlled reasons, and combinations of these goals and reasons, known as goal complexes. In linear regression analyses, with sets of items intended to represent separate constructs, self-based goals (R^2 adjusted = .120, $p < .001$), autonomous reasons (R^2 adjusted = .117, $p < .001$), and task-based goals (R^2 adjusted = .105, $p < .001$) were predictors of TOEIC Listening scores.

However, two models of goal complexes (Sommet & Elliot, 2017) that combined sets of items intended to represent mastery and performance goals with autonomous reasons, rather than individual sets of items, accounted for a greater share of the variance (21–24%). Controlled reasons for goals, and related goal complexes, did not predict TOEIC L & R scores. However, the project was limited by the small number of items per construct ($k = 3$) and by item quality.

The aim of the current study is to improve item quality for items that are intended to represent autonomous reasons for goals and approach goals from the 3×2 model. Rasch analysis will be used to investigate item quality.

Methodology

Setting

This project is situated at SU, a public university in central Japan. The university has approximately 250 students in each of Years 1 and 2 where the English program was

concentrated. The university is composed of two faculties, neither of which are English. The largest faculty, Business Management, a pseudonym, comprises 70% of the students. The remaining students were in Health Education, also a pseudonym. Nearly two-thirds of the students are female; and approximately half of the student population comes from outside the prefecture where the university is located. All students participate in required overseas study in their second or third year depending on faculty.

TOEIC L&R was used as a KPI for the English language programme (ELP). For Cohorts 2018–2022, all students were required to reach 600 points or higher. Per faculty, attainment rates were 70% for Business Management and 30% for Health Education. For cohorts entering from 2023, the KPI were separated into three different benchmarks and was different between faculties. The new TOEIC L&R KPIs for Business Management and Health Education respectively are: (A) average: 730 and 650 points; (B) growth rates: 40% and 30%; and (C) the average of the top 20%: 800 and 700 points.

To evaluate the ELP in regard to the above KPI, students complete the TOEIC L&R four times, in early April and late January of Year 1, and mid-November and late January of Year 2. Students have four 100-minute required English classes per week in Year 1: two classes are in an Accuracy stream and are taught through English and Japanese; two classes are in a Fluency stream and are taught in English. In Year 2, students have between two and four required classes per week depending on the faculty. In Year 1 Fluency classes, students must also demonstrate mastery of the NGSL (Browne, 2013) as part of their grade. That is, the NGSL test accounts for 10% of the course grade. For study purposes at the university, the 2801-word NGSL list is subdivided into five levels by frequency, Levels 1 and 2 are each composed of 700 words, Levels 3 and 4 are each composed of 500 words, and Level 5 is composed of the remaining 401 words.

In April of Year 1, students complete a paper-based receptive—word-meaning recognition—levels check. Items were adapted from Bennett and Stoeckel (2013). The levels check covers levels 1–3 only, and a score of 90% or higher was required to pass a level. Students' starting level was the lowest level they did not pass, or Level 4 for students who pass all three tested levels. Students were then assigned study materials based on their level, and at the end of the quarter completed an NGSL test composed of both receptive and productive items (Richard, 2018). For students who passed the end-of-quarter test, they received 10 class points and advanced to the next level the following quarter. For students who did not pass the end-of-quarter test, they received a score from 0-to-7.5 class points and stayed at the same level the following quarter. Students who pass Level 5 advance to the New Academic Word List (NAWL).

Instrument and Participants

A 26-item questionnaire was prepared using Microsoft Forms. The L1-Japanese questionnaire was composed of 5-item sets of Likert-scale items for autonomous reasons for goals, task-based goals, self-based goals, and other-based goals, the 20 variables that are the focus of this study. These items were adapted from (Richard, 2025) which had been based on Sommet and Elliot (2017) and Elliot et al (2011). The questionnaire also included a demographic question, three long-form written questions addressing reasons for goals, goals, and other comments, and finally two consent questions, addressing use of questionnaire and test data. Only the results from the Likert-scale items are discussed in this paper (see appendix).

Three versions of the above questionnaire were created with different reference points of goals and reasons for goals. One version had 20 items referencing NGSL; a second version had 20 items referencing TOEIC L&R; and a third version had 20 items each referencing NGSL and TOEIC L&R. All Year 1 students randomly received by email one of the three versions on the day they completed their end-of-Quarter 2 NGSL test. Of these, 40% of Year 1 students completed the questionnaire, with 30 students each completing the two shorter versions and 36 students completing the longer version. All 96 of these students gave consent for their data to be used in the study. For research at SU involving test scores and questionnaires, consent alone is sufficient. Scores per NGSL level of the study participants were compared with the non-participants. At each level, average scores were within a $\pm 2\%$ band, indicating that on average, the students who responded to the questionnaires had similar average abilities as those who did not respond. Independent-samples *t*-tests indicated no significant differences in mean NGSL scores between questionnaire respondents and non-respondents at any NGSL level ($p > .05$), suggesting comparable ability distributions.

Analyses

Rasch analysis was used to investigate construct validity using guidelines by Fisher (2007). See Table 1. Probability curves were also inspected to detect irregular usage patterns, and Andrich thresholds were inspected for orderly advancement, with at least 1.4 logits, but no more than 5.0 logits, between thresholds. All analyses were done using Winsteps, Version 4.4.5 (Linacre, 2006).

Table 1

Fisher's (2007) Guidelines for Construct Validity

	Poor	Fair	Good	Very Good	Excellent
Fit Mean Square	< .33 - > 3.0	.34 - 2.9	.5 - 2.0	.71 - 1.4	.77 - 1.3
Person & Item Reliability	< .67	.67 - .80	.81 - .90	.91 - .94	> .94
Person & Item Strata	2 or less	2-3	3-4	4-5	>5
Ceiling (max extreme scores)	> 5%	2-5%	1-2%	.5-1%	<.5%
Floor (min extreme scores)	> 5%	2-5%	1-2%	.5-1%	<.5%
1st Contrast Unexplained Variance	> 15%	10-15%	5-10%	3-5%	<3%

Results

As a first step, dimensionality for all 20 items was investigated. The percentage of unexplained variance; that is, variance which was not accounted for by the Rasch model, in the first contrast was greater than 10% ($EV = 5.98$; 11.7%). This process was repeated for items intended to represent achievement goals ($k = 15$; $EV = 6.39$; 14.2%), and mastery-based goals—task- and self-based goals—combined with autonomous reasons ($k = 15$; $EV = 4.64$; 12.2%). In all three cases, the raw unexplained variance in the first contrast indicated that the items were likely multidimensional. No further analyses were repeated for the above combinations of items. This process was also repeated for the 10 items intended to represent mastery-based goals. The results suggested these items were potentially unidimensional ($k = 10$; $EV = 2.57$; 9.3%). This result was consistent with prior findings (see Richard, 2022). Thus, the following sets of items were further analyzed using Rasch analysis: autonomous reasons for NGSL Goals (AR, $k = 5$), mastery-based NGSL goals (MG, $k = 10$), and other-based goals (OG, $k = 5$).

Autonomous Reasons for Goals ($k = 5$)

The five items intended to represent AR were investigated using Rasch analysis. Referring to Fisher's (2007) guidelines, person and item reliability and strata, which was the number of statistically distinct performance levels the test can differentiate, were good, as were infit. Infit mean square (MNSQ) indicates how well item responses

match model expectations, with a focus on unexpected responses to items targeted to the person's ability. However, Andrich thresholds were disordered, and the percentage of unexplained variance in the first contrast was greater than 10% (18.1%) (see the appendix). In addition, 9% of the participants had extreme scores (maximum measures). In a second run, categories were collapsed (123456→ 333456), which improved the Andrich thresholds; however, the percentage of unexplained variance in the first contrast remained high (16.5%). In a third run, one item, AR2, "the reason I have NGSL goals is because I enjoy the challenge", was temporarily deleted. Unfortunately, the percentage of unexplained variance in the first contrast remained high (15.1%). No combination of deleting one or two items improved the percentage of unexplained variance in the first contrast.

Mastery-Based NGSL Goals ($k = 10$)

The items intended to measure task-based goals ($k = 5$) and self-based goals ($k = 5$) were analyzed together, as mastery-based goals, using Rasch analysis. In an initial run, person reliability and strata were very good, but item reliability and strata were poor, and Andrich thresholds were disordered (see appendix). In addition, 18% of the participants had extreme scores (maximum measures). In a second run, categories were collapsed (123456→ 333456), which improved the Andrich thresholds; however, item reliability and strata remained poor. Deleting combinations of 1 or 2 items improved item reliability and strata (although they remained poor according to Fisher's (2007) guidelines; however, the percent of unexplained variance in the first contrast increased.

Other-Based NGSL Goals ($k = 5$)

The items intended to measure other-based goals ($k = 5$) were analyzed using Rasch analysis. In an initial run, person reliability and strata were good, but item reliability and strata were poor, and Andrich thresholds advanced by more than 5.0 (see the appendix). In addition, 15% of the participants had extreme scores (maximum measures). In a second run, categories were collapsed (123456→ 333456), which improved the Andrich thresholds; however, item reliability and strata remained poor. Deleting combinations of 1 or 2 items improved item reliability and strata, although they remained poor according to Fisher's (2007) guidelines.

Discussion

Three sets of items, autonomous reasons for NGSL goals (AR, $k = 5$), mastery-based goals (MG, $k = 10$), and other-based goals (OG, $k = 5$), were analyzed using the Rasch model. For each set of items, most, but not all of the guidelines from Fisher (2007) were met, especially after collapsing categories. For AR, the primary concern was the high percentage of unexplained variance in the first contrast. However, these five items, while assumed to represent autonomous reasons for NGSL goals, are in fact composed of items intended to represent intrinsic motivation; for example, "The reason I have goals related to NGSL study is because I enjoy the challenge (AR2), and items intended to represent integrated regulation" + for example, "The reason I have goals related to NGSL study is because I think it is worth it" (AR4). Thus, it might not be surprising that the percentage of unexplained variance in the first contrast was found to be high.

For both MG and OG, the primary concern was the low item reliability and strata. For both sets of items, there was a ceiling effect, that is many respondents achieved the highest possible score, limiting variability, for a large percentage of participants. Items were written with the intention of separating students into different ability levels, such as "My goal is to get a better grade than the last time I took this test" (SG1) and My goal is to do as well as possible compared to the last time I took this test (SG3); however, participants saw little difference in difficulty between these two items (0.24 logits and -0.18 logits). This is likely due to the fact that at SU, students receive a pass or fail, not a score. The results suggest that items need to be refined to target higher ends of the logit scale. In Rasch analysis, the logit scale is the log-odds unit that represents person ability and item difficulty on the same interval scale. Examples of refined items might be "My goal is to get the highest score possible", or simply "My goal is to master the NGSL."

Students at SU are expected to master the NGSL in their first year at the university. NGSL study and testing is part of classroom-based activities, homework assignments, and class grades. Knowing this, it is possible that the goal of NGSL mastery has been integrated into participants' study systems by many students. I am also the faculty member in charge of the NGSL at the university. Knowing this, students might have been reluctant to disagree with the NGSL goal-related statements. Even though the Microsoft Form that students responded to indicated that answering the questionnaire, or not, had no effect on their grades, many students might have been anxious about being perceived as not interested in NGSL mastery.

One ultimate goal of this paper is to refine a set of items to develop combinations of reasons and goals—goal complexes—for NGSL study and the TOEIC L&R. The results identified AR, MG, and OG items that performed better than others. These can be used

to generate goal complex items. Data for this new study will be gathered in late January 2025. In addition, after one year of study in the English programme at SU, it has been found that approximately 15-20% of the students do not demonstrate mastery of the NGSL. An immediate need is to report on the relationships between NGSL levels and test scores with goals and reasons for their goals. This is currently being investigated and will be reported in 2025. Furthermore, when responding to the questionnaires, most participants also include extended written comments about their goals and reasons for those goals. These written responses need to be investigated and used to triangulate questionnaire responses. Finally, one long-term goal is to develop and improve materials for students based on clusters of goals and reasons for goals. This is happening internally and is ongoing.

Conclusion

The study described in this paper took place at SU, a public university with high TOEIC-related KPIs and where students must demonstrate mastery of the NGSL. The goal of this paper was to describe the validation of questionnaire items related to autonomous reasons for study and achievement goals. These items will eventually be used to identify clusters of students, with specific goals and reasons for their goals, and to provide specific study materials. In the study, various forms of questionnaires composed of autonomous reasons for NGSL goals ($k = 5$), mastery-based goals ($k = 10$), and other-based goals ($k = 5$) were randomly assigned to Year 1 students at SU. Approximately 40% of the students responded, with no difference in their average vocabulary ability between students who answered the questionnaire and those who did not. Rasch analysis was used to investigate the items. These three sets of items performed well on most, but not all, of the guidelines described by Fisher (2007). Although potential reasons for this were described, the results indicate that some items need further revision.

Bio Data:

Jean-Pierre J. Richard, EdD, is an associate professor in the Faculty of Global Management at the University of Nagano. His research focuses on vocabulary acquisition, individual differences, and assessment. He is currently developing NGSL vocabulary learning materials for students with lower proficiency levels, is leading a collaborative textbook project for academic discussion, and is investigating how goal complexes influence student engagement with vocabulary learning and standardized language testing. richard.jean-pierre@u-nagano.ac.jp

References

- Bennett, P., & Stoeckel, T. (2013). Developing equivalent forms of a test of general and academic vocabulary. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings*. JALT.
- Browne, C. (July, 2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 37(4), 13-15.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-519. <https://doi.org/10.1037/0022-3514.80.3.501>
- Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 x 2 achievement goal model. *Journal of Educational Psychology*, 103(3), 632-648. <https://doi.org/10.1037/a0023952>
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transaction*, 21(1), 1095. <https://www.rasch.org/rmt/rmt211m.htm>
- Linacre, J. M. (2006). WINSTEPS Rasch measurement computer program (Version 4.4.5) [Computer software]. Chicago. <https://www.winsteps.com/>
- Richard, J.-P. J. (2018). Developing Written Productive Items for an Institution-Level NGSL Assessment Program. *Vocabulary Education & Research Bulletin*, 7(2), 14-20.
- Richard, J.-P. J. (2022, June 23-25). *Initial Validation of a Japanese-Language 3 x 2 Achievement Goal Questionnaire for (L2-English) Vocabulary Study* [Poster presentation]. PLL4, Cape Breton University, Sydney, Nova Scotia, Canada. <https://doi.org/10.13140/RG.2.2.26221.56804>
- Richard, J.-P. J. (2025). Goals, Reasons, and Goal Complexes as Predictors of TOEIC Listening Scores. *The Global Management*, 12, 1-21. <https://doi.org/10.32288/0002000130>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sommet, N. & Elliot, A. (2017). Achievement goals, reasons for goal pursuits, and achievement goal complexes as predictors of beneficial outcomes: Is the influence of goals reducible to reasons? *Journal of Educational Psychology*, 109(8), 1141-1162. <https://doi.org/10.1037/edu0000199>
- Sommet, N., Elliot, A. J., Sheldon, K. (2021). Achievement goal complexes: Integrating the “what” and the “why” of achievement motivation. In R. W. Robin, & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (4th ed., pp. 104-121). Guilford Press.

Appendix

Rasch Results Based on Fisher (2007)

	Reliability		Strata		Infit	Outfit	PMCs	Raw Unexplained Variance (First Contrast)		Andrich Threshold Orderly Advance	
	Person	Item	Person	Item				Eigenvalue	Percent	>1.4	<5.0
Autonomous Reasons for NGSL Goals (k = 5)											
AR (k = 5)	.83	.95	3.33	6.03	.79 - 1.43	.73 - 1.34	.76 - .87	2.61	18.1%	No	Yes
AR (k = 5) ¹	.83	.95	3.28	6.41	.81 - 1.27	.78 - 1.26	.78 - .87	2.52	16.5%	Yes	Yes
AR (k = 4) ²	.80	.96	2.99	6.18	.79 - 1.16	.76 - 1.13	.83 - .85	1.97	15.1%	Yes	Yes
¹ Collapsed categories: 123456 → 333456											
² Collapsed categories and deleted AR2 (I enjoy the challenge)											
Mastery-Oriented NGSL Goals (k = 10)											
MG (k = 10)	.91	.13	4.48	0.85	.67 - 1.62	.52 - 1.59	.76 - .88	2.53	9.3%	No	Yes
MG (k = 10) ³	.72	.41	2.45	1.44	.57 - 1.63	.35 - 1.89	.65 - .82	2.67	15.4%	Yes	Yes
MG (k = 9) ⁴	.70	.40	1.43	1.07	.57 - 1.68	.42 - 2.15	.64 - .86	2.40	15.3%	Yes	Yes
MG (k = 8) ⁵	.70	.57	2.39	1.89	.63 - 1.45	.61 - 1.42	.72 - .90	2.61	20.3%	Yes	Yes
³ Collapsed categories: 123456 → 333456											
⁴ Collapsed categories and deleted TG1 (My NGSL goal is to get many questions right)											
⁵ Collapsed categories and deleted TG1 and TG2 (My NGSL goal is to answer questions correctly)											
Other-Based NGSL Goals (k = 5)											
OG (k = 5)	.82	.31	3.20	1.23	.41 - 1.64	.39 - 1.33	.90 - .94	1.59	5.1%	Yes	No
OG (k = 5) ⁶	.91	.31	4.67	1.23	.41 - 1.61	.42 - 1.38	.90 - .96	1.60	6.9%	Yes	Yes
OG (k = 4) ⁷	.90	.61	4.23	1.99	.58 - 1.34	.51 - 1.16	.93 - .97	1.62	8.1%	Yes	No

⁶ Collapsed categories: 123456 → 333456

⁷ Collapsed categories and deleted OG1 (*My NGSL goal is to perform better than other students*)