

# 外国人指導助手(ALT)の教師的役割を形成する英語教師の振る舞い—中学校のチームティーチング授業を対象とした会話分析研究—

## Promoting Assistant Language Teachers' Roles in Team-Teaching Classrooms: A Conversation Analytic Study in Japanese Junior High Schools

石野未架(いしの みか)

Mika Ishino

同志社大学

*Doshisha University*

近年の学校英語教育では、英語を母語とする外国人指導助手 (Assistant Language Teacher : ALT) と、日本語を母語とする英語教師 (Japanese Teacher of English : JTE) が共に授業を行うチームティーチングが推奨されている (文科省, 2016)。しかし実際の授業ではALTの役割が軽視されるという問題が指摘されてきた (Tajino & Tajino, 2000)。本研究ではこの問題を背景に、授業におけるJTEのどのような振る舞いがALTの役割を軽視、または重視することにつながるのかを明らかにする。中学校におけるチームティーチング授業の15時間分の録画データを会話分析により検証した結果、ALTに評価の機会を与えないというJTEの振る舞いがALTの役割を軽視することにつながる実践として観察された。また反対に、ALTに生徒への評価を委ねるといったJTEの振る舞いがALTの役割の重要性を演出することが明らかになった。会話分析により記述されたこのようなJTEの実践は、他のJTEがチームティーチング授業で方略的に活用できる実践知である。

To enhance the internationalization of school education, the Japanese Ministry of Education is hiring native English speakers as assistant language teachers (ALTs) (MEXT, 2016). ALTs are expected to provide students with authentic English-speaking resources in team-teaching classes with Japanese teachers of English

<https://doi.org/10.37546/JALTJ43.1-3>

*JALT Journal, Vol. 43, No. 1, May 2021*

(JTEs). However, ALTs' underestimated role in these team-teaching classes has been reported as a substantial problem. ALTs have frequently been called "human tape recorders" (Mahoney, 2004), as they are often only required to play the role of a model speaker in the class. Although many studies point to the problem of ALTs not being included in the classroom (Takeda, 2017) and suggest that their lack of awareness as teachers in team-teaching classrooms is caused by JTE practices during the classes, no practical solutions have yet been suggested for appropriate teacher trainings based on empirical research. Therefore, this study will examine which kinds of JTE practices work for promoting or demoting the ALT's role in team-teaching classrooms.

In order to analyze teacher practices during team-teaching classes in detail, many previous studies have employed conversation analysis (CA) (Aline & Hosoda, 2006; Lee, 2016; Park, 2014). Those studies have demonstrated the effectiveness of the method for understanding the complex interactional practices of the teachers in team-teaching classes. Thus, this study also employed CA as a useful application for examining detailed JTE practices during team-teaching classes with ALTs. Using CA, the researcher conducted a comparative case study on 15 team-teaching lessons of two pairs of team teachers (each pair consisting of an ALT and a JTE) in two public junior high schools. Using these data from two pairs of teachers, an analytic focus was placed on how these teachers shared their turn-takings in "Initiation-Responses-Evaluation (IRE) sequences" (Mehan, 1979), which have been frequently analyzed in previous studies (e.g., Lee, 2016) as an analytic framework for classroom interactions.

The analysis revealed that when one of the ALTs took over the floor for evaluation, it made a significant impact by promoting and raising the ALT's role as a teacher in the team-teaching lesson. Even for cases in which the JTE (the ALT's teaching partner) initiated IRE sequences by taking turns for evaluation, the ALT promoted his or her own role in the lesson. Interestingly, in all cases, the ALT took turns for evaluation, only after initiation by the JTE. In contrast, because the ALT in other cases did not take turns for evaluations or initiations of IRE sequence, the ALT's role was underestimated, and was viewed as a "human tape recorder" in the classroom. The ALT's loss of opportunity for evaluation was caused by the JTE's taking over the former's turn for evaluation. These findings show that a key practice for promoting the role of ALTs in the team-teaching classrooms is for JTEs to hand over the evaluation to ALTs.

The findings of this study suggest that teachers taking turns for evaluation in an IRE sequence helps promote the significance of the teachers' roles in the classroom. This is particularly important given the complex classroom setting of team-teaching in Japan. Providing turns for ALTs to conduct evaluations in IRE sequences would be useful practical knowledge for JTEs to organize better team-teaching classes. Although this exploratory study only observed two pairs of team-teaching lessons, the findings of this study will contribute to providing a practical guide for future trainings for team teaching in Japan.

**Keywords:** 会話分析; ティームティーチング; IRE連鎖; 教師教育; 教室インタラクション; conversation analysis; team-teaching; IRE sequence; teacher education; classroom interaction

近年の学校英語教育では、文部科学省(以降、文科省)が発表した「グローバル化に対応した英語教育改革実施計画」(文科省, 2016) 等を受けて、英語を母語とする外国人指導助手(Assistant Language Teacher:ALT)の雇用が推奨されている。ALTの役割は日本語を母語とする英語教師(Japanese Teacher of English:JTE)とティームティーチングを行い、英語圏の文化や言語表現を教えることである(e.g., MEXT & The British Council, 2013; 文科省, 2013)。

しかし実際のティームティーチング授業では、ALTにその役割を発揮する機会が十分に与えられてないことが指摘されている(e.g., Takeda, 2017)。例えば、ALTの授業における役割が実際には「人間テーブルレコーダー」になっているという指摘は20年程前から報告されている(Mahoney, 2004; Tajino & Tajino, 2000)。ALTの人間テーブルレコーダーの役割とは、授業では主にJTEが依頼した場面で特定の英語のモデル音声を生徒に披露するという役割である。人間テーブルレコーダーとしてのALTには、生徒の質問に直接答えたり、生徒のパフォーマンスを評価したりするような「教師的行為」(岡田, 2015)を行使する機会が与えられない。このように教師的行為を行使する機会が与えられないALTは、教室で教師としてのアイデンティティを形成しにくいことが分かっている(奥貫・ルイス, 2012; Tsuido et al, 2012)。そして教室で教師としてのアイデンティティを形成しにくいという経験は、ALTの学校現場への適応困難や、中途退職の要因となることが報告されている(Carless, 2006; 石野, 2018)。

ALTとJTEのティームティーチングを有効に運用するためには、ALTが教師としてのアイデンティティを形成することができる機会を授業で保障していく必要がある。そのためには、JTEがALTに教師的行為を行使する機会を与えるよう意識的に振る舞うことが重要である。多くの中学・高等学校ではJTEが主導でティームティーチングを展開しており、加えて生徒と母語を共有している。授業中に生徒に何か問題があればJTEはすぐに生徒の母語を使用して対処することができる。したがってJTEの授業における振る舞いは、ALTが授業で教師的行為を行使する機会を得られるか否かに大きく影響を与えるのである(e.g., Ishino, 2018)。

では、具体的にJTEのどのような振る舞いがALTに教師的行為を行使する機会を与えるのだろうか。JTEのティームティーチングに対する心的態度を調査した研究は多く存在する一方(e.g., Ogawa, 2011; 田邊, 2010)、JTEのティームティーチングにおける具体的な振る舞いに焦点をあてた研究は僅かである(Aline & Hosoda, 2006)。今後の学校英語教育で益々導入が見込まれるALTとのティームティーチング授業を有効に展開するためには、JTEのどのような振る舞いがALTの教師的役割を形成するのか、あるいは形成しないのかについて実際のデータに基づき検証する必要があるだろう。

そこで本研究では、ティームティーチング場面の先駆的研究であるAline and Hosoda(2006)らの手法を足掛かりとし、実際のティームティーチング授業でALTの役割形成に影響を及ぼすJTEの振る舞いを検証する。Aline and Hosoda(2006)らは、会話分析(Sacks, 1992; Schegloff, 2007)の手法を用いてティームティーチングにおける

教師らの振る舞いを微視的に分析した。会話分析は教室における教師と生徒の微細なやりとりを分析するのに有効な手法であり、近年の言語教育研究において広く用いられてつある。本研究でもそのような会話分析の手法を用いて、チームティーチングにおけるJTEのどのような振る舞いがALTの教師的な役割の形成を阻害、または促進するのかを分析する。そして、ALTの教師的役割を形成するJTEの具体的な振る舞いを明らかにすることで、チームティーチング授業を有効に運用するための提言を行う。

### 教室場面の会話分析研究とチームティーチング授業

チームティーチング授業における教師の振る舞いを会話分析の視点から分析する試みはAline and Hosoda(2006)らの研究を皮切りに韓国を中心に広がりつつある(Park, 2014; Lee, 2016)。これらの研究において基盤となったのは、教師1人対生徒多数という通常の授業場面を対象とした会話分析の知見である。次項ではまず、このような通常の教室場面における会話分析の知見を概観する。

#### 教室場面の基本的会話構造:IRE連鎖

教室場面を対象とした会話分析研究の重要な知見として、IRF/E連鎖(McHoul, 1978; Mehan 1979a; Sinclair & Coulthard, 1975)がある。IRF/E連鎖とは、教師による応答の促し(Initiation)－生徒による応答(Response)－教師によるフィードバック(Feedback)または評価(Evaluation)という行為の連鎖のことである。便宜上、本稿ではIRE連鎖を使用する。

以下の抜粋1はIRE連鎖の説明でよく用いられるMehan(1979b)のデータである。

抜粋1 (Mehan(1979b, p.285)より引用)

- 1 T: what time is it, denise?
- 2 S: 2:30
- 3 T: very good, denise

1行目では教師が生徒であるDeniseに今何時なのかという質問を開始する。2行目ではこの質問への応答として、Deniseが2:30と答えている。そして3行目ではDeniseの応答について教師がvery goodと評価をしている。Mehanはこのような三つの順番で構成される会話の連鎖を日常会話のそれと比較したのである。日常会話では、第三の位置には評価がくるのではなくthank you deniseなどといった感謝の言葉が最も観察される。したがって、応答の促し(I部分)-応答(R部分)-評価(E部分)という教師と生徒によるIRE連鎖は教室場面に特有の行為連鎖なのである(Mehan, 1979a, 1979b, McHoul, 1978)。

このようなIRE連鎖は、教室場面を対象とする会話分析研究の基盤となってきた。これまでの研究では、IRE連鎖のI部分を教師がいかにデザインするのか(e.g., Okada, 2010)、IRE連鎖の適切な位置で生徒がいかに新たな連鎖を開始するのか(e.g.,

Waring, 2009)など、IRE連鎖を基盤に教師や生徒の様々な振る舞いが明らかにされてきた。

しかし、先行研究が一人の教師が授業を運営する場면을対象としてきたのに対し、本研究では2人の教師が授業を運営するチームティーチングの場면을対象としている。このため、IRE連鎖における教師の振る舞いやその組織化はより複雑になる。次項では、そのようなチームティーチングの授業場面において2人の教師がどのように振る舞うのか、どのようにIRE連鎖を展開するのかという点に焦点をあて、先行研究を概観する。

### チームティーチング授業における教師の振る舞い

チームティーチング場면을対象にした会話分析研究では、Aline and Hosoda(2006)らの研究が先駆的なものとして知られている。Alineらは日本の小学校で学級担任(homeroom teacher:HT)とALTがチームティーチング授業でそれぞれどのように振る舞うのかを分析した。そしてHTの授業への参加の方法を次の4つのパターンにわかれている。それは、1)傍観者、2)翻訳者、3)共同学習者、4)共同授業者である。AlineらはHTの授業への参加パターンとALTの教師的な役割の関係については分析していないが、チームティーチングの複雑な相互行為の分析において会話分析が有効であることを示した点でこの分野に大きな影響を与えた。

Park(2014)は、韓国の小学校における外国人の英語指導助手(ALT)と韓国人英語教師(Korean English Teacher: KET)のチームティーチング授業を対象に、教師や生徒の授業における参加構造がどのように組織されるのかを分析した。そして主に次の2つの構造が観察されることを報告した。1つ目の構造は、IRE連鎖を主導する教師と連鎖における第二位置を取得する生徒、そして生徒たちが第二位置を取得しやすいようジェスチャーなどでその補助をする教師の3者間の構造である。まずどちらかの教師が第一の発言順番を取得してIRE連鎖を開始する。そして生徒らが第二位置で発言順番を取得しやすいように一方の教師がジェスチャーなどを用いて生徒らに何らかのヒントを示す。ヒントを得て生徒らが第二位置で発言順番を取得するという構造である。2つ目の構造は、一方の教師がIRE連鎖を主導し、もう一方の教師がその第二位置を取得して、生徒らは「傍観者」(Aline & Hosoda, 2006)として相互行為に参加するという構造である。この構造では2人の教師の英語によるやりとりが展開され、生徒らはそのやりとりを傍観するかたちとなる。

Lee(2016)もまた、Park(2014)の発見した2つ目の構造と同様の構造を報告している。Lee(2016)が韓国の小学校においてALTとKETのチームティーチングを分析した事例では、ALTが開始したIRE連鎖の第二位置でKETが発言順番を取得してモデル回答を披露し、ALTがそれを評価するという実践が観察された。Leeはこのような教師間の即興的なやりとりが、生徒らにとって自然な英会話に触れる機会になると主張し、その教育的価値を主張した。これら先行研究は、チームティーチングという複雑な授業の構造を整理し、理解するうえで重要な知見を導いている。

しかし先行研究では、IRE連鎖のなかで授業を主導する教師の振る舞いがもう一方の教師の役割形成にどのように影響を与えるのかについては関心を示してこなかった。本研究では、IRE連鎖において授業を主導するJTEのどのような振る舞いがALTの教師的役割の形成に影響を与えるのかという点に関心をよせ、JTE主導のチームティ

ーチング場面を分析対象とする。そして以下の課題を設定することで、ALTの脱テープレコーダー化を図り、教師としての役割形成を促すJTEの具体的な振る舞いを明らかにする。

研究課題： ティームティーチング場面においてJTEのどのような振る舞いがALTの教師的役割を形成、または抑制するのだろうか。

### 方法及びデータの概要

本研究では、先行研究 (Park, 2014; Lee, 2016) にならい、会話分析の方法を用いて実際のティームティーチング授業の録画データを分析する。観察されたIRE連鎖のすべてを分析範囲とし、IRE連鎖を構築する過程でJTEのどのような振る舞いがALTの教室における役割を構築するのか、またはしないのかという点を分析の焦点とする。具体的にはまず、先行研究で問題視されてきたALTの人間テープレコーダー化が観察される事例を対象に、人間テープレコーダー化を導いたと考えられるJTEの振る舞いを分析する。次に、人間テープレコーダー化の事例とは反対にALTが教師的行為を行使する事例を対象に、そこでのJTEの振る舞いを分析する。そしてそれぞれの事例の比較から、ALTの脱テープレコーダー化を促し、教師的な役割を形成するJTEの具体的な振る舞いを明らかにする。

分析に使用したデータは、関西の中規模の公立中学校2校(それぞれA校、B校とする)で収集したティームティーチングの授業録画データ15時間分である。A校、B校とも収集したのは1年生の英語のティームティーチング場面である。対象とした授業はそれぞれ2012年から2016年にかけてA校およびB校にて撮影されたものである。A校のティームティーチングはJTEのヤマザキとALTのジュディ、B校ではJTEのナカムラとALTのブラウンで行われた(全て仮名)。

なお、分析の節で示す抜粋は全て会話分析で一般的に用いられるJefferson(2004)による文字化の原則(付録参照)を用いて書き起こしたものである。

### 分析

まず、JTEの振る舞いがALTを人間テープレコーダー化してしまう事例として、A校のヤマザキとジュディのティームティーチング場面を分析する。次に、JTEの振る舞いがALTに教師的行為を行使する機会を保障する事例として、B校のナカムラとブラウンのティームティーチング場面を分析する。

### ALTを人間テープレコーダー化するJTEの振る舞い

抜粋1はA校1年A組のティームティーチング授業が開始される場面である。この授業では授業の開始時に、生徒ら全員が起立して教師に礼をするという儀礼的やりとりが行われていた。1行目のジュディの発話は、生徒らが礼を終えて着席した直後に開始されたものである。

抜粋2(ヤマザキ=Y、ジュディ=J、生徒I=S1、複数の生徒=Ss)

- 1 J: ↑good morning every↑one  
 2 (0.9)  
 3 Ss: good morning.miss|[Ju:dy:  
 4 Ss: [Yamaza[ki  
 5 Ss: [Judy  
 |((ジュディは頷きながら生徒に笑顔をむける))  
 6 (0.3)  
 7 Y: |hhhh(h)ちや(h)う(h)ちや(h)う(h)=  
 |((Yは笑いながら右手を顔の前で左右に振ってみせる))  
 8 Y: =|ジュディ先生がいてくれるねん(h)か(h)ら(h)  
 |((両手でジュディのほうを指し示す))  
 9 (0.3)  
 10 Ss: Ju:|dy:(.)[(a:nd)  
 |((ジュディは自分を指さし生徒のほうに笑顔を向ける))  
 11 Y: [ >one more time<(.)もういっかいいきま:す。  
 12 (0.6)  
 13 J: ↑good morning every↑one

抜粋2は、授業開始場面(e.g., Mehan, 1979a)に観察される典型的なIRE連鎖のようにみえる。しかし、7行目、そして11行目、18行目のヤマザキの発話順番は、これまで報告されてきたティームティーチングにおけるIRE連鎖の構造(Park, 2014; Lee, 2016)からは逸脱したものとなっている。これまで観察されてきたIRE連鎖では、I部分を産出した教師がそのままE部分を産出して連鎖を終了させるというやりとりが観察されてきたが、この抜粋ではI部分の産出者とE部分の産出者が異なっている。

まず、7行目でJTEのヤマザキがALTのジュディが開始した挨拶(I部分)の連鎖への生徒の反応(R部分)について「ちやうちやう」と否定的評価(E部分)を与えている。続く8行目では挨拶のあてさきをジュディに訂正するよう生徒らに促している。一方で、1行目でジュディが開始した挨拶に対する生徒の反応(R部分)をジュディがどのように扱っているのかは5行目で彼女が頷きながら生徒に笑顔を向けていることから観察できる。ジュディの頷くという非言語的な行為は、3～5行目の生徒のR部分に肯定的反応を示しており、少なくとも修正の必要性を示すものではないことがわかる。したがってここではジュディのR部分に対する評価とはおおよそ異なる評価が連鎖の開始者ではないヤマザキによって行われ、生徒らへの訂正が促されたことになる。

同様の現象は10行目と11行目でも観察される。10行目ではジュディが笑顔を向けて自分のほうを指さしてみせることで、進行中の生徒の訂正の実行を受け入れていることが観察できる。しかし、11行目でヤマザキが生徒らの訂正の実行途中である10行

目の発話に重ねてone more timeと再度やりなおすよう促している。ジュディによって開始されたIRE連鎖のE部分はヤマザキによってひきとられ、修復の開始もヤマザキによって促されているのである。つまり、ジュディが生徒の間違いを訂正したり、評価したりするといった教師的行為を行う機会は、ヤマザキの振る舞いによって奪われているといえるだろう。

加えて、11行目の「もういっかいいきます」というヤマザキの発話も注目すべき組み立てである。11行目のヤマザキの発話は生徒に顔が向けられたまま産出されているが、0.6秒の沈黙の後、ジュディが再びgood morning everyoneと挨拶の連鎖を開始している。このことから、11行目のヤマザキの連鎖の開始の促しは、生徒だけではなくジュディに挨拶の開始を求めるものとしても受け止められたことがわかる。ヤマザキによるジュディへの挨拶の開始の依頼や承諾といった相互行為 (Park, 2014) が実演されることなく、ジュディは連鎖の開始を行っている。通常期待されるはずの依頼行為が省かれているため、ジュディの拒否や承諾といった応答の機会も省かれていることになる。したがって、ここでのジュディは、ヤマザキの開始の合図によって自動的に挨拶のやり直しを行う存在であり、依頼や承諾の相互行為を実演する必要のない存在としてあつかわれていることがわかる。

このようにALTから評価の機会を奪ってしまうJTEの振る舞いや、ALTへの相互行為的配慮を示さないJTEの振る舞いは、まさにALTを人間テープレコーダー化してしまう振る舞いといえるのではないだろうか。本節では、このようなJTEの振る舞いがALTのテープレコーダー化を促す振る舞いと捉え、同様の事例(抜粋3~4)を更に分析し、検証していく。

以下に示す抜粋3は、抜粋2と同じA校1年A組のティームティーチング授業である。ここでは、教師らが生徒にHow many X do you have?という英語表現を教えている。黒板にはHow many X do you have?という英文が書かれ、黒板にむかって右手にヤマザキ、そして左手にジュディが立っている。

抜粋3(ヤマザキ=Y、ジュディ=J、生徒1=S1、複数の生徒=Ss)

- 1 Y: 猫は::何匹かっていますか::って聞きたいときに:(.)ちよっと、
- 2 ジュディ先生に教えてもらいたいと思いま:す
- 3 (2.3)
- 4 Y: |はい  
|((黒板のhow many cats do you haveという英文を指さしながら))
- 5 J: how many ↑cats do you |ha:ve.  
|((生徒のほうをみながら))
- 6 (1.6)
- 7 Ss: <ho::w many: ca::ts do you: |ha:ve>  
|((ジュディうなずく))
- 8 S1: <how [many>
- 9 Y: [はいそいいいま:す(.)]



10 S1: [(dog)

11 Y: [はい

注目したいのは、1行目から3行目のヤマザキの発話の組み立てである。1行目から3行目の発話は全て生徒のほうを見ながら産出されているため、ヤマザキがこの発話の受け手を生徒に設定して産出していることがわかる。続く4行目のヤマザキの「はい」という発話とともにヤマザキが目線を送っているのも生徒である。したがって1行目から4行目までのヤマザキの発話は生徒を受け手として産出されているといえる。しかし、4行目に続いて発言順番を取得しているのはジュディである。ジュディはヤマザキが指し示している黒板の英文をそのまま読み上げるかたちで発話を産出している。したがってヤマザキの1行目から3行目、そして5行目の発話は、生徒を受け手に志向された組み立てであると同時に、ジュディにモデル発話を行うよう促されたものとしてジュディに捉えられたことが観察できる。

ここでも抜粋2と同様に、ヤマザキとジュディの間には教師間の依頼や承諾といった相互行為(Park, 2014)が実演されていない。例えば4行目でヤマザキは「はい」という発話と共にジュディに視線を送ることもできるわけである。けれどもそうとはせずに、生徒らに視線を送りながら片手で黒板に書かれたhow many?の英文指し示している。ヤマザキからジュディへのモデル音声披露の依頼という通常期待されるはずの行為が省略されているために、ジュディによる拒否や承諾といった応答の機会も省かれることになる。ジュディにはモデル発話の産出を拒否する相互行為上の機会が与えられていない。したがって、ジュディの授業における相互行為への「参与」(Goffman, 1981)はヤマザキのそれと比較して極めて限定的である。

しかしまた、このようなヤマザキによるジュディへの依頼の省略、ジュディの限定的な授業への参与が、ジュディには非日常のものとして扱われてはいないことがわかる。何故なら、4行目から5行目への発話順番の交代が何の問題もなく行われているからである。仮に、ジュディがヤマザキによる依頼の省略を依頼の不在として扱うのであれば、4行目のヤマザキの発話の後にはジュディの何らかの行為によって彼女がそう捉えたことが示されるはずである。例えば、ヤマザキによる依頼が存在しないために、5行目ではジュディが何も発話をしないということが考えられる。その場合はヤマザキがジュディにモデル発話の依頼を明示的に行うという「連鎖の拡張」(Schegloff, 2007)が観察されるだろう。しかし、実際にはヤマザキとジュディの間でこのようなやりとりは発生せず、4行目と5行目の発話順番の交代が素早く行われている。したがって、このようなジュディとヤマザキのやりとりが授業ではこれまでに繰り返し行われてきたものと解釈できる(e.g., Sacks, 1985)。

次に注目すべき特徴は、7行目でジュディが生徒たちの英語発話に非言語的に示した「うなづく」という評価行為が参与者に評価行為として扱われていないという点である。6行目のジュディの発話はヤマザキの促しにより開始されたものではあるが、ジュディの5行目の発話を受けて生徒らが7行目でそれを繰り返しているので、ジュディの5行目の発話はIRE連鎖のI部分とも捉えることができる。そして8行目の生徒らの発話の終わりにジュディが示した「うなづく」という行為は、IRE連鎖のE部分として捉えられる。しかし、ヤマザキが8行目のジュディの非言語的な評価行為に続いて10行目で言語的な評価を行うのである。そしてそのまま12行目で一連の連鎖を終了させて

いる。つまり、ジュディが生徒の発音を評価するという教師的行為を行ったにも関わらず、ヤマザキが評価を上書きし、そのままIRE連鎖を終了させたのである。したがって、このようなヤマザキの一連の振る舞いはジュディが評価という教師的な行為を行わせる機会を奪うものだといえるだろう。

このようにジュディから評価するという教師的行為の機会を奪ってしまうヤマザキの振る舞い(9行目)や、ジュディを相互行為的配慮の必要のない存在として扱ってしまうヤマザキの振る舞い(4行目)は頻繁に観察された。以下の抜粋4でもまた、そのようなヤマザキの振る舞いを観察することができる。

抜粋4は、抜粋3の授業の続きであり、教師らがhow many Xという英語の質問に対してどのように答えるべきかを生徒に教えている場面である。抜粋4の1行目は、ヤマザキが複数の犬の絵が描かれたカードを指さしながら生徒のほうに顔をむけて開始した発話である。

抜粋4(ヤマザキ=Y、ジュディ=J、複数の生徒=Ss)

- 1 Y: はい(.) |ジュディ先生にきいてもらいます  
| ((Yは犬が複数描かれたカードを指しながら))
- 2 (0.6)
- 3 J: how many dogs do you have?
- 4 (0.3)
- 5 Ss: I have a [four do:g
- 6 Ss: [three [dogs
- 7 Ss: [many dogs
- 8 (0.4)
- 9 Y: many dogs. (0.2) はい、たくさんいますね、
- 10 (0.6)
- 11 Y: はい、|え::(.)じゃあ、はい(.)ねこさんは?  
| ((Yは犬のカードを猫が3匹かかれたカードに切りかえる))
- 12 (0.2)
- 13 J: how many cats do you have?
- 14 Ss: I have [three cats
- 15 Ss: [(how tree cats)
- 16 Y: how t(h)r(h)e(h)e(h) catsなんていま(h)せ(h)ん(h)よ=
- 17 Y: =誰です(h)か(h)ハウトゥリーキャッツとかいうてる(h)の(h)=
- 18 Ss: =|hahahahahahahahaha
- 19 | ((ジュディは首をかしげてみせる))
- 20 Y: はい、

まず確認できるのは、1行目と3行目、及び11行目と13行目におけるヤマザキによるジュディへの連鎖開始の依頼の省略である。1行目及び11行目でヤマザキはいずれも生徒のほうに顔を向けたまま発話を産出しているため、それらの発話が生徒に向けられていることがわかる。しかし、続く3行目や13行目でジュディが英語による質問の連鎖を開始していることから、それらが同時にジュディへの連鎖開始の指示となっていることもわかる。つまりここでもジュディへの連鎖開始の依頼という相互行為的な配慮は示されないのである。

次に確認できるのは、9行目および16行目から20行目のヤマザキの発話によって、ジュディの評価の機会が奪われているという点である。まず、ジュディが開始した質問のIRE連鎖に5行目から7行目で生徒らが答えているが、これらR部分に対応するE部分は、ヤマザキによって9行目で産出され連鎖が終了している。また、13行目からジュディが開始した質問に対応する生徒らの14行目から15行目のR部分は、16行目から17行目でヤマザキによって評価され、20行目でヤマザキによって連鎖の終了が示されている。したがって、ここでもヤマザキの振る舞いによって評価というジュディの教師的な行為の機会が奪われているのである。

以上の分析から特に重要なのは、授業場面のIRE連鎖において連鎖を開始した教師がE部分を産出する機会をもたないという点である。これは、これまでのチームティーチングを対象とした研究(Park, 2014; Lee, 2016)では報告されてこなかったものである。これまでIRE連鎖におけるE部分は教師の権力性を最も発揮する部分とされ、教室会話の特徴として論じられてきた行為である(e.g., Mehan, 1979b)。したがって、ジュディが取得することが期待される発言順番の位置でその発言機会を奪い、代わりに評価を行うというヤマザキの振る舞いは、ジュディの教師的役割の形成を阻害する振る舞いだといえるだろう。このように、教室におけるALTの教師的役割形成においては評価の機会を与えられるか与えられないかという点が特に重要になるのではないだろうか。

そこで次項では、これまでの事例とは対照にIRE連鎖のなかでJTEが評価の機会をALTにゆだねるという行為が観察された事例を分析する。B校で観察されたナカムラとブラウンのチームティーチング授業から、JTEがALTに評価の機会をゆだねることによってALTの教師的役割がどのように形成されるのかを検証する。

### ALTの教師としての役割を形成するJTEの振る舞い

抜粋5は、B校1年B組の授業でALTのブラウンが出身地であるカナダの食べ物を生徒に紹介している場面である。複数の生徒らがブラウンの紹介した食べ物に英語で言及するが、1行目ではブラウンが生徒らのハンバーガーの発音について間違いを指摘している。そこで3行目からはJTEのナカムラが生徒らに発音訂正に挑戦するよう連鎖を開始している。

抜粋5(ブラウン=B、ナカムラ=N、生徒1=S1、生徒2=S2、複数の生徒=Ss)

- 1 B: hannbargar じゃなくて hamburger
- 2 (0.8)
- 3 N: おう(0.3)やってみようぜ

- 4 (0.4)
- 5 S1: よし.hhh いけ
- 6 (0.2)
- 7 N: ん?うちら(.)>うちらうちらうちら<(.)
- 8 : まず-(.)まずうちらがやってみよう
- 9 (1.0)
- 10 B: hamburger
- 11 (0.2)
- 12 S1: humberger
- 13 (0.7)
- 14 S2: humberger=
- 15 Ss: =hahahahahahahahaha
- 16 N: おう>うまいうまい<(. )いまのだけ?
- 17 S2: こいつ(. )こいつ
- 18 N: もっかいやってみろ.
- 19 (0.4)
- 20 S1: humberger:
- 21 (0.6)
- 22 N: |どう?  
| ((NはBをみながら))
- 23 | (0.4)  
| ((BはNをみる))
- 24 S1: humberger:
- 25 B: hamburger:
- 26 S1: hamberger:
- 27 B: okay. that's pretty |close.  
| ((Bは二回うなずく))

注目すべきは、22行目でナカムラがブラウンへ評価機会を譲渡している点である。ナカムラは生徒1のハンバーガーという発音に関して既に16行目で「うまいうまい」と肯定的評価を行っている。しかしナカムラはここで連鎖を終了することなく、18行目で生徒に再度発音するよう促している。そして生徒が20行目で再度披露した発音には評価を行わずにブラウンにその評価の機会をゆだねているのである。さらに、ブラウンも23行目ですぐに評価を行うことはせず、生徒1が24行目でターンを取得して再度披

露した発音についてもすぐには評価を与えていない。25行目で自らがモデル発音を披露して連鎖を拡張し、発音の訂正を促している。そして生徒1がブラウンの発音を真似て繰り返すという訂正の応答を産出した後、ようやく評価を行っているのである。

ここでのブラウンの評価の組み立ては、ナカムラの評価と比較すると異なる特徴もっている。ブラウンは、16行目でナカムラが行った「うまいうまい」という評価に対して、that's pretty closeという定式をとっている。ナカムラが行った「うまい」という評価は発音の上手い・下手という技術的な尺度における評価である。これに対してブラウンのpretty closeという評価の定式では、ブラウンが構音した発音を基準に、それに近いのか・遠いのかという尺度における評価を行っている。したがって発音を産出する技術の得手不得手を軸に評価を行ったナカムラに対し、ブラウンは自身のモデル音声にどれだけ近いのか遠いのかという評価軸に志向しているといえる。ブラウンがもしナカムラの評価軸に寄り添い生徒の発音技術の得手不得手の観点から評価を行うのであれば、27行目ではthat's pretty goodという定式をとることもできたわけである。しかしここでブラウンが産出した評価はthat's pretty closeであったことから、ブラウンはナカムラとは異なる評価軸を明確に使用して生徒1の発音の評価を行ったと解釈できる。

つまり、一連の連鎖で観察されたブラウンの評価行為は英語母語話者である自分の発音を評価軸として用いた評価行為なのである。そしてこの行為は、ナカムラの評価の機会をゆだねるという振る舞いによって行使する機会が与えられたものである。ブラウンによって開始された発音の訂正を促す連鎖は、ナカムラの参加によって拡張し、更にナカムラが評価の機会をゆだねることによって、ブラウンが評価を行い終了したのである。ここで観察されたブラウンの行為は、抜粋1で観察されたジュディの人間テープレコーダーとしての行為とは異なり、生徒の間違いの訂正を促したり生徒の発音を評価したりするという教師的行為であった。そしてその評価には英語母語話者であるALT自らが基準として用いられていた。したがって、JTEではなくALTにしか果たすことのできない教師としての役割がここでは形成されたといえるだろう。

このようにALTの教師的役割を形成する、評価をゆだねるというJTEの振る舞いについて、更に3つの事例(抜粋6~8)を検証していく。抜粋6は同じく1年B組の授業で、ブラウンが出身地カナダに関する情報を英語で述べた後、生徒らが英語で聞き取れた情報を教室の皆と共有する活動場面である。ここでは3行目で生徒1が言及したカナダに棲息するクマの品種grizzly bearの望ましい発音をめぐるやりとりが行われている。

抜粋6 (ブラウン=B、ナカムラ=N、生徒1=S1、生徒2=S2、複数の生徒=Ss)

- 1 S1: >はいはい<は:はい
- 2 | (0.8)
- 3 | ((NはS1を指名する))
- 3 S1: crispy:: bears(.)か:なんか
- 4 Ss: おお::
- 5 (0.6)

- 6 N: 発音もつかいな  
 7 S1: crisby bear=  
 8 S2: =gu.hh gu.hh=  
 9 S1: =ku.hh ku.hh cris[bee::  
 10 Ss: [hahahahahahahaha  
 11 N: >おうおう.<おまえらがまず修正せ:,  
 12 (1.0)  
 13 S2: greas bee bear  
 14 N: おう(.)hhhちやう.皆にいけ  
 15 S2: GREASILY::BEAR=  
 16 Ss: =hahahahahahahahaha  
 17 N: ちょ.hhみな笑わんと  
 18 (0.4)  
 19 N: もっかいこ:(.)>はい<  
 20 Ss: greasily:: bear=  
 21 N: = greasily bear  
 22 Ss: hahahahahahahahaha  
 23 N: Is that brown sensei okay?  
 24 |(0.2)  
 |((Bは2回うなずく))  
 25 B: grizzly bear (.) that's pretty close  
 26 N: |grizzly bear  
 |((N2回うなずく))

注目したいのは、やはりここでも抜粋5と同様にナカムラがブラウンに評価の機会をゆだねている点である。しかし抜粋5と異なるのは、ナカムラが開始したIRE連鎖の第三位置(21行目)で評価を行わずに、自らも評価を受けるべき行為を産出したうえで、評価の機会をブラウンにゆだねている点である。抜粋5では、ナカムラが生徒の発音の披露に対して16行目で「おう>うまいうまい<」と評価を行ったのに対し、ここでは、20行目の生徒の発音の披露につづけて21行目で自らもgreasily bearと発音を披露している。そしてそのうえでブラウンに評価行為を求めることによって、ここでのブラウンが評価すべき対象に自らの発音も含めているのである。

そしてブラウンはここでも再度自らモデル発音を披露したうえでthat's pretty closeという評価の定式をとっている。実際にそれまで生徒やナカムラが発音していたものはgreasily bear(/grɪ:sɪli beə/)であり、25行目でブラウンが披露した発音のgrizzly

bear (/grizli béə/)とは異なるものであった。これらが参加者にとっても理解可能な差異として受け取られていることが26行目のナカムラの自己訂正から観察することができる。つまり25行目のブラウンの評価行為は、むしろ訂正を促すものとしても観察可能である。そしてその訂正の促しは生徒だけではなくナカムラにも向けられたものとして扱われていることがわかる。ここでも、ナカムラの評価をゆだねるという行為がブラウンの教師としての役割を形成しているといえる。加えて、ナカムラ自らの発音も対象とした評価の機会をブラウンにゆだねることで、ブラウンにはナカムラと生徒の両方の発音を評価することが出来る者としての役割が形成されたのである。

以上、抜粋5と6から観察されたブラウンの教師としての役割は、モデル音声を披露するだけではなく、それに基づいた発音の評価を行うという役割であった。そしてこのようなブラウンの役割は、ナカムラの評価をゆだねるという行為によって形成されることがわかった。そしてこのようなALTの教師としての役割は、発音の評価に限って形成されるわけではないことも他の事例によって確認された。それらの事例を次の抜粋7及び8でとりあげる。

抜粋7は、B校1年C組の授業でナカムラが生徒らに「スキー講習を受けた」という英語表現について生徒にこたえさせる連鎖を開始した場面である。1行目では、生徒が ski lessoned と回答したのに対してナカムラが第三位置を産出している。

抜粋7(ブラウン=B、ナカムラ=N、生徒1=S1、生徒2=S2、生徒3=S3、複数の生徒=Ss)

- 1 N: 残念ながら:こうゆう単語はな-(0.4)ないんですよ?  
 2 (0.2)  
 3 B: sorry:: (0.2) [no::  
 4 S1: [took?  
 5 N: [a ski lesson (.)の[まえ  
 6 S2: [あ::tookや=  
 7 S1: =°took°  
 8 (1.2)  
 9 N: took やつたらいけますか?=  
 10 B: =that's right (.) yeah:: took is a good word ( )  
 11 S3: |took か::  
 |((Bは黒板に何かを書いている))  
 12 Ss: took  
 13 (0.2)  
 14 N: あるいは::(0.2) ha::d is okay?  
 15 (0.3)  
 16 N: had (.) ha::d=

- 17 B: =hard?(.) h↑ad? Eichi ay dee had?(.)  
 18 B: y↑eah (.) we had=  
 19 N: =we ha:d a ski lesson  
 20 (0.2)  
 21 N: うん  
 22 Ss: had

注目したいのは、ナカムラが1行目と9行目で生徒のR部分に対する評価をブラウンにゆだねた行為、そして14行目で拡張連鎖を開始した行為である。まず1行目でナカムラは生徒が答えたski lessonedという表現について、そのような表現は存在しないと宣言するようにみえたが、発話の産出途中でブラウンにその正誤の判断をゆだねている。つまり、生徒のR部分に対して構築するE部分(ここでは訂正)の産出を途中でブラウンにゆだねているのである。ブラウンは3行目の連鎖位置でゆだねられた訂正を即座に行い、教師的役割を果たしている。更にその後、複数の生徒からtookというR部分が産出されるが、ナカムラはこれらを評価することなく9行目でブラウンにtookでも可能かどうかと尋ね、評価の機会をゆだねているのである。そしてブラウンがthat's rightと10行目でE部分を産出することで教師的行為を行っている。

加えて、13行目で終了しかけた「スキー講習を受けた」という英語表現をめぐる一連のIRE連鎖について、14行目からナカムラは自らの質問によって拡張連鎖を開始している。ここでナカムラは、英語表現に関して自らがブラウンに質問をすることで「共同学習者」(Aline & Hosoda, 2006)として振る舞っているのである。そして、学習者として振る舞うことによって、ブラウンにこの場において適切な英語表現を教えることができる者としての「知識権限」を譲渡している(Raymond & Heritage, 2006)。つまり、ここではナカムラの知識権限の譲渡によって、ブラウンが英語の専門家として振る舞う機会が形成されているのである。

以上のように、ナカムラの質問によって拡張された一連の連鎖や、ナカムラが評価行為をゆだねることによって構築された連鎖位置には、ブラウンに教師的な役割を果たす機会が設けられている。そして同時にナカムラが評価行為をゆだねることで、ナカムラには評価することが出来ない生徒の応答(R部分)について評価(E部分)することができるブラウンの「知識の権限領域」(Raymond & Heritage, 2006)が形成されているのである。

次の抜粋8の授業は1年C組の授業が録画された後、休み時間をはさんだ次の時間に行われたD組での授業である。抜粋7で観察された「スキー講習を受けた」という英語表現をめぐるやりとりは、D組の授業でも同様に観察された。

抜粋8(教師ブラウン=B、教師ナカムラ=N、生徒1=S1)

- 1 N: we?  
 2 S1: received=  
 3 N: =received (.) なるほど::



- 4 (0.2)
- 5 N: received (.) とかな::  
( (11行分省略) )
- 17 N: can we use this received?
- 18 B: we::ll it's.hh-(0.9) a:: it's a little::  
19 (0.3)
- 20 N: (かたい?)=
- 21 B: =そ:う little(.)[strong
- 22 N: [ (かたい) okay (.) okay
- 23 : do you have a good word for this?
- 24 B: ah:mm:(.) yeah ai-(.) I'd like(0.2) to say(.) ahm:(.)  
25 we(0.2) took a ski lesson=
- 26 N: =okay

1行目でナカムラは、weを主語にしてski lessonを目的語にした際にweの後に来る動詞を生徒にこたえさせるべく不完全化発話(Koshik, 2002)を産出している。これを受けて2行目では生徒1がreceivedと答えている。6行目からは別の生徒が授業とは関係のない活動をしてナカムラがそれを叱るという連鎖が展開されたため、その部分は省略している。

注目したいのは、17行目でナカムラがreceivedという生徒の語彙の選択について自らの評価を曖昧にし、その評価をブラウンにゆだねた点である。3行目と5行目でナカムラは生徒の応答によって提案されたreceived (a ski lesson) という表現について「なるほど」と反応し、receivedという表現について否定的な評価を下してはいないが0.2秒の沈黙の後に「receivedとかな」と発話し、まだ他の表現の可能性を示している。そして17行目でナカムラはブラウンにreceivedの使用可能性についての判断を求めるのである。ここでもナカムラは、英語の言語知識に関わる評価の産出機会をブラウンにゆだねることによって、ブラウンの知識の権限領域を形成し、ALTにしか果たすことのできない役割を演出しているといえる。

## 考察

これまでJTEのどのような振る舞いがALTの教師的役割の形成に影響を与えるのかについて、7事例のIRE連鎖を分析してきた。最初の3つの事例(抜粋2から4)では、ALTを人間テーブルレコーダー化するJTEの特徴的な振る舞いを分析した。次の4つの事例(抜粋5から8)では、人間テーブルレコーダー化とは反対にALTの教師的役割の形成を促すJTEの振る舞いを分析した。その結果、前者と後者ではIRE連鎖のE部分でJTEがどのように振る舞うのかに大きな違いがあることが明らかになった。

まず、抜粋2から4で観察されたティームティーチング授業では、ALTはIRE連鎖を開始(I部分の構築)する機会は得ることができても、生徒たちの反応(R部分)に

ついて、評価やフィードバックを与える(E部分を産出する)機会を得ることはなかった。ALTにE部分を産出することが期待される連鎖位置で、JTEがE部分を産出するという場面が3つの抜粋全てにおいて観察された。つまり、ALTがIRE連鎖の中でE部分を産出するという教師的行為を行使する機会は、全てJTEによって奪われていたのである。このようにIRE連鎖の開始者ではない教師がE部分を産出するというやりとりは、先行研究(Park, 2014; Lee, 2016)では観察されなかった相互行為である。更にもう1点、先行研究とは異なる特徴として、教師間の依頼や承諾といった即興的なやりとりが全てにおいて観察されなかった点も挙げられる。ALTはJTEから依頼や承諾といった相互行為的配慮が示されないままモデル音声の披露を行い、それに反応した生徒への評価(E部分)の機会は、JTEによって奪われていた。このようなIRE連鎖におけるALTの限定的な参与とその役割は、まさに人間テープレコーダーとしてとらえられるものだろう。

一方で、抜粋5から8で観察されたチームティーチングでは、IRE連鎖においてE部分の産出がJTEによりALTにゆだねられるという実践が観察された。評価をゆだねるという実践により演出されたのは、英語の発音や表現といった特定の教授領域におけるALTの知識権限(Heritage & Raymond, 2005)と知識権限領域(Raymond & Heritage, 2006)であった。この演出において、ナカムラがブラウンに評価をゆだねる行為と、ブラウンがナカムラから評価の機会を受け取るという行為はそれぞれの教師の知識権限や知識権限領域に結びついた合理的行為として生徒らに理解可能になる。特定の領域(抜粋5と6では発音、抜粋7と8では言語表現)の評価をブラウンにゆだねるというナカムラの行為は、教師的な役割を果たすことのできるALTという存在を演出する振る舞いだともいえる。つまり、特定の場面でJTEがALTに評価をゆだねるという行為はALTの専門性を演出し、ALTの教室における教師としての価値を形成していたのである。

以上の結果から、IRE連鎖においてJTEがALTの評価の機会を奪うのは、ALTの教師的役割の形成を阻害する振る舞いであることがわかった。一方で、JTEがALTに評価の機会をゆだねるという振る舞いは、JTEがALTの教師的役割を形成するために授業で行える具体的な実践であることがわかった。

しかしここで忘れてはならないのは、ALTにとって評価という行為があくまでもJTEにゆだねられることによって実行可能な行為だという点である。ALTが評価という行為を実行するにあたっては、JTEからその実行機会をゆだねられるという相互行為が実演されなければならないのである。つまり、JTEから局所的に知識権限を譲渡されるという意味において、ALTに形成される役割はあくまでも「指導助手」としての「準教師的」な役割なのだといえよう。

## 結論

本研究では、日本国内の中等教育学校で行われた外国人指導助手(ALT)と日本人英語教師(JTE)によるチームティーチング場면을対象に、IRE連鎖のなかで形成されるALTの役割を会話分析により検証した。ALTの役割が軽視され、人間テープレコーダー(e.g., Mahoney, 2004)化されるという問題を背景に、JTEのどのような振る舞いがALTの教師的役割の形成を阻害、または促進するのかという問いをたて、2組のJTEとALTのチームティーチング場面を比較検証した。

分析の結果、ALTの教師的役割の形成には、IRE連鎖においてJTEがALTにE部分(評価の機会)の産出をゆだねるという振る舞いが重要であることがわかった。また、反対にIRE連鎖においてALTに評価を行う機会を与えないというJTEの振る舞いがALTの教師としての役割形成を阻害することがわかった。加えて、ALTの人間テープレコーダー化はモデル音声の依頼と承諾の省略という相互行為によって構成されることがわかった。本研究で得られた知見は、ティームティーチングにおいてJTEがALTの教師的役割を形成するために実践できる具体的な振る舞いとして、今後の教師教育に貢献できる知見である。また、ALTの人間テープレコーダー化という事象が具体的にどのような相互行為によって構成されるのかを明確にした点も、会話分析により導かれた重要な知見である。

本研究で記述されたJTEの振る舞いは、会話分析の視角から自然主義的に記述された相互行為上の実践である。したがって記述された実践は、他の教師も今後の授業に活用することが可能である。既に日本国内の先行研究(岡田, 2015)が示しているように、教師の複雑な教育実践の記述にとって会話分析は非常に有効な研究手法なのである。本研究では授業のIRE連鎖に限定して分析を行ったが、IRE連鎖以外の相互行為に観察されるJTEとALTの社会文化的な関わりを分析することで、新たな授業実践に関わる知見を得られる可能性もある。そのようなIRE連鎖の枠組みにとられないティームティーチング授業の分析は、今後の研究課題としていきたい。

## 謝辞

本研究はJSPS科研費 JP19K13305の助成を受けたものである。

## 引用文献

- Aline, D., & Hosoda, Y. (2006). Team teaching participation patterns of homeroom teachers in English activities classes in Japanese public elementary schools. *JALT Journal*, 28, 5-21.
- Carless, D. R. (2006). Good practices in team teaching in Japan, South Korea and Hong Kong. *System*, 34, 341-351.
- Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.
- Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*, 68, 15-38.
- 石野未架(2018)「退職を選んだ外国人指導助手 (ALT) が語るティームティーチングの課題: アクティブ・インタビューを用いて。」『言語文化共同研究プロジェクト』2017, 33-42.
- Ishino, M. (2018). Micro-longitudinal conversation analysis in examining co-teachers' reflection-in-action. *System*, 78, 130-147.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In: G. H. Lerner (Ed.), *Conversation analysis: Studies from the first Generation* (pp. 13-31). John Benjamins.

- Koshik, I. (2002). A conversation analytic study of yes/no questions which convey reversed polarity assertions. *Journal of Pragmatics*, 34, 1851–1877.
- Lee, J. (2016). Teacher entries into second turn positions: IRFs in collaborative teaching. *Journal of Pragmatics*, 95, 1–15.
- Mahoney, S. (2004). Role controversy among team teachers in the JET program. *JALT Journal*, 26, 223–244.
- McHoul, A. (1978). The organization of turns at formal talk in the classroom. *Language in Society*, 7(2), 183–213.
- Mehan, H. (1979a). *Learning lessons*. Harvard University Press.
- Mehan, H. (1979b). 'What time is it, Denise?': Asking known information questions in classroom discourse. *Theory into Practice*, 18(4), 285–294.
- 文部科学省 (2013)『グローバル化に対応した英語教育改革実施計画』<[http://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/\\_icsFiles/afieldfile/2014/01/31/1343704\\_01.pdf](http://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2014/01/31/1343704_01.pdf)> (閲覧日: 2017年6月30日)
- Ministry of Education, Culture, Sports, Science and Technology, The British Council. (2013). ALT handbook. Retrieved from: [https://www.britishcouncil.jp/sites/default/files/alt-handbook-en\\_0.pdf](https://www.britishcouncil.jp/sites/default/files/alt-handbook-en_0.pdf) [Accessed 30 December, 2019].
- 文部科学省 (2016)『教育課程部会 外国語ワーキンググループ(第5回)配付資料』『資料7初等中等教育の英語教育の推進に係る取組』<[http://www.mext.go.jp/b\\_menu/shingi/chukyo/chukyo3/058/siryu/\\_icsFiles/afieldfile/2016/01/15/1366027\\_10.pdf](http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/058/siryu/_icsFiles/afieldfile/2016/01/15/1366027_10.pdf)> (閲覧日: 2017年5月)
- Ogawa, C. (2011). Perceptions about team teaching: From assistant language teachers and Japanese teachers of English. In A. Stewart (Ed.), *JALT2010 Conference Proceedings*. Tokyo: JALT.
- Okada, Y. (2010). Repairing “failed” question in foreign language classrooms. *JALT Journal*, 32(1), 55–74.
- 岡田悠佑 (2015)「アイデンティティによる尺度化: 言語教師の定式化手続きの会話分析研究」『*JALT Journal*』37, 147–170.
- 奥貫妃文・ルイス, カーレット (2012)「労働者としての ALT (外国語指導助手) についての一考察: 公教育の非正規化, 外注化の観点から」『大阪経済法科大学アジア太平洋研究センター年報』9, 17–24.
- Park, J. E. (2014). English co-teaching and teacher collaboration: A micro-interactional perspective. *System*, 44, 34–44.
- Raymond, G., & Heritage, J. (2006). The epistemics of social relations: Owning grandchildren. *Language in Society*, 35, 677–705.
- Sacks, H. (1992). *Lectures on conversation, Volume I & II*. Blackwell.

- Sacks, H. (1985). The inference-making machine: Notes on observability. In T. van Dijk, (Ed.), *Handbook of discourse analysis, Vol. 3: Discourse and dialogue* (pp. 13–23). Academic Press.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1). Cambridge University Press.
- Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Tajino, A., & Tajino, Y. (2000). Native and non-native: What can they offer? Lessons from team-teaching in Japan. *ELT Journal*, 54, 3–11.
- Takeda, Y. (2017). More than a tape-recorder: negotiations of English language teacher identities in the JET Program in Japan (Doctoral dissertation, University of British Columbia).
- Tsuido, K., Otani, M., & Davies, W. (2012). An analysis of assistant language teachers' perceptions of their working relationships with Japanese teachers of English, *広島外国語教育研究*, 15, 49–64.
- Waring, H. Z. (2009). Moving out of IRF (Initiation-Response-Feedback): A single case analysis. *Language Learning*, 59(4), 796–824.

### 付録 <トランスクリプト記号一覧>

---

(0.2)	0.2秒の音声のない状態,
(.)	短い0.2秒以下の音声のない状態
=	途切れなく密着した発話
[	発話の重なりの開始
-	言葉が不完全で途切れた状態
:	直前の音の引き伸ばし
↑	直後のイントネーションの上昇
h	呼気音
> <	周辺よりも速い発話
< >	周辺よりも遅い発話
<u>あ</u>	周辺よりも協調された発話
((あ))	観察された非言語情報の記述

---

# Research Forum

## Evaluating a Scale's Construct Validity to Assess the Group Work Environment Using the Rasch Model

Mitsuko Tanaka  
*Osaka University*

Tanaka (2017) developed a questionnaire instrument comprised of 2 constructs (*group cohesion* and *group engagement*) to assess the group work environment using exploratory factor analysis. The purpose of this study was to examine the construct validity of the scale using Rasch analysis. The sample included 200 second-year Japanese university students who engaged in group work over a semester. The results of the Rasch analysis verified construct validity with acceptable fit statistics, adequately high reliability and separation indices, logical hierarchical ranking of the items, and the unidimensionality of the construct.

Tanaka (2017) は探索的因子分析を用いて2つの構成概念(「グループの結束性」と「グループの積極的な関与」)を抽出し、グループワーク環境を測定する質問紙を作成した。本研究の目的はラッシュモデルを用いて当該尺度の構成概念妥当性を検証することである。本研究の調査対象者は一学期にわたりグループワークを行った日本の大学二年生200名である。ラッシュモデルによる分析の結果、適合度統計値が許容範囲内に収まること、信頼性係数と分離指数が十分に高いこと、項目の階層序列が論理的であること、構成概念の一元性が確保されることが明らかになり、尺度の構成概念妥当性が支持された。

**Keywords:** group work; learning environment; Rasch analysis; scale validation

<https://doi.org/10.37546/JALTJ43.1-4>

*JALT Journal*, Vol. 43, No. 1, May 2021

Learners' motivation to study a second language (L2) tends to increase in a positive classroom environment and decrease in a negative environment (Chang, 2010). As motivation is a major determinant of learning behaviors (Dörnyei & Ushioda, 2013), the nature of the learning environment either directly or indirectly influences L2 learning achievement (Kozaki & Ross, 2011; Sasaki et al., 2017). Although the learning context's significance is well-acknowledged in the field of L2 acquisition (cf. Dörnyei & Ushioda, 2013), there has been little quantitative investigation into the contextual effects on L2 learning motivation in small group work settings. Consequently, instruments that assess the group work environment are extremely limited. This study focused on Tanaka's (2017) questionnaire instrument to measure the group work environment and examined its construct validity using the Rasch model (Rasch, 1980).

### Group Work Environment and Motivation

In his three-level framework outlining L2 motivation, Dörnyei (1994) listed *group cohesion* as one of the group-specific components influencing motivation at the level of the learning situation. Clément et al. (1994) incorporated the concept of *group cohesion* or *group dynamics* and demonstrated the importance of the social dimension (i.e., the learning environment) for understanding the motivation of L2 learners. They developed eight questionnaire items to measure the construct of group cohesion and demonstrated its adequately high reliability (Cronbach's  $\alpha = .77$ ) as well as some associations with L2 learning-related factors. Their scale has been employed to demonstrate the impact of group cohesion on self-efficacy and autonomy (Chang, 2010) and language production (Dörnyei & Kormos, 2000) at the whole-class level.

As L2 classroom learning generally entails social interactions, group dynamics have mostly been the primary focus of research concerning learners' motivation involving the social dimension of the L2 classroom (e.g., Dörnyei & Murphey, 2003). However, researchers have investigated different aspects of the classroom learning environment, such as the normative classroom environment operationalized by students' perceptions of their classmates' career pursuits (Kozaki & Ross, 2011; Sasaki et al., 2017) and the classroom's social climate (Joe et al., 2017). Research focusing on the small group work setting has also examined diverse environmental properties, including group work dynamics (Poupore, 2016, 2018), group-directed motivational currents (DMCs; Dörnyei et al., 2016), group vision (Dörnyei & Kubanyiova, 2014), and collective-efficacy (Leeming, 2020). As a learning environment

can entail various aspects, researchers have explored the different facets according to the specific purposes of their studies.

As already mentioned, instruments to measure the group work environment are extremely limited. There appear to be only three scales designed to assess group properties in the field of L2 learning motivation. Poupore (2016) developed an instrument to measure group work dynamics without relying on self-reporting (i.e., questionnaires). His scale was intended to measure group work dynamics based on observations of both verbal and nonverbal behaviors in peer interactions. Using the scale, he demonstrated positive associations between group work dynamics, motivation, and language production both quantitatively (Poupore, 2016) and qualitatively (Poupore, 2018). Leeming (2020) created a questionnaire scale to assess *collective-efficacy* for group work in English communication classes and demonstrated how collective-efficacy evolves over time through L2 group work. Tanaka (2017) developed a questionnaire instrument to measure the group work environment that comprises two constructs: *group cohesion* and *group engagement*. Using the scale, Tanaka (2018) revealed the significant impact of the group work environment on motivation regardless of learners' English proficiency level. Although Tanaka (2017) developed and validated the scale using exploratory factor analysis, "[Rasch measurement theory can] play an important role in the process of construct validation, in that a set of test or questionnaire items constitute the instrument designer's empirical definition of the construct" (Sick, 2008a, p. 3). An evaluation of the scale using Rasch analysis would therefore add further evidence in demonstration of the construct's validity. Accordingly, the purpose of this study is to examine the construct validity of the scale developed by Tanaka (2017) using Rasch analysis. The following research question was posited in this study:

- RQ. Does Tanaka's (2017) questionnaire instrument designed to assess the group work environment in fact measure what it purports to?

## Method

### Participants

This study used the data collected by Tanaka (2018) for her study in the department of sports and health science at a private university in Japan. The participants were 200 second-year Japanese undergraduates (118 males and 82 females) who voluntarily agreed to take part in the study and com-



pleted the questionnaire. They were enrolled in a project-based learning (PBL) course comprising 15 total sessions (90 minutes per session) over a semester. They stayed in the same group, consisting of a maximum of six members, and performed two group projects (i.e., a debate and panel discussion) during the semester. Example topics for the debates and panel discussions were “Should children have mobile phones?” and “How to become one of the world’s top tennis players,” respectively. They delivered three group presentations (i.e., a debate presentation, as well as midterm and final panel discussion) in total, and wrote one group paper based on a final panel discussion. At the end of the semester, they responded to a questionnaire that included items about the group work environment. They had varying levels of English proficiency, based on their TOEIC scores ( $M = 418.69$ ,  $SD = 111.64$ ). Although they did not particularly enjoy learning English in the PBL setting, as their level of intrinsic motivation was neither high nor low, they were not discouraged.

## Instrument

This study evaluated a questionnaire scale developed by Tanaka (2017) for the purpose of assessing the group work environment. Tanaka (2017) created items in Japanese, drawing on the classroom climate inventory (CCI; Ito & Matsui, 2001) using a 6-point Likert scale (1 = *Strongly disagree* to 6 = *Strongly agree*). Using exploratory factor analysis to identify underlying relationships among the data collected from participants, two factors were extracted: Group Cohesion ( $k = 6$ ) and Group Engagement ( $k = 6$ ). Both factors exhibited high Cronbach’s  $\alpha$  values (.91 and .85 for Group Cohesion and Group Engagement, respectively).

Prior to the Rasch analysis, data screening was conducted based on Tabachnick and Fidell (2007). No missing data were found in the data set. Three univariate and three multivariate outliers were identified, based on the criterion of 3.29 in standardized scores ( $p < .001$ , two-tailed test), and the critical value of chi-square:  $\chi^2(12) = 32.909$  at  $p < .001$ . As those six outliers accounted for less than 5% of 200 students, they were retained with further analysis. Table 1 shows descriptive statistics of all 12 items. It should be noted that no univariate and multivariate outliers were identified based on person measures for two constructs (i.e., Group Cohesion and Group Engagement).

**Table 1**  
*Descriptive Statistics of 12 Items*

	<i>M</i>	<i>SE</i>	95% CI		<i>SD</i>
			LB	UB	
<b><i>Group Cohesion (COHE)</i></b>					
COHE1	4.49	.08	4.33	4.64	1.14
COHE2	4.58	.07	4.43	4.72	1.02
COHE3	4.69	.07	4.54	4.83	1.01
COHE4	4.14	.08	3.98	4.29	1.08
COHE5	4.59	.07	4.45	4.72	1.00
COHE6	4.11	.08	3.94	4.27	1.18
<b><i>Group Engagement (ENGA)</i></b>					
ENGA1	3.95	.08	3.80	4.10	1.07
ENGA2	4.10	.08	3.95	4.25	1.09
ENGA3	3.61	.08	3.45	3.77	1.14
ENGA4	3.92	.07	3.78	4.06	0.99
ENGA5	3.90	.07	3.75	4.04	1.05
ENGA6	4.61	.07	4.46	4.75	1.02

*Note.* Students responded using a 6-point Likert scale (1 = *Strongly disagree* to 6 = *Strongly agree*). CI = confidence interval.

## Data Analysis

This study employed the Rasch-Andrich rating scale model (Andrich, 1978) using Linacre's (2013) Winsteps computer program (Version 3.80.0). Although the Rasch model (Rasch, 1980) assumes dichotomous data (e.g., right or wrong), a Likert scale has more than two response choices (e.g., "strongly disagree," "disagree," "agree," and "strongly agree"). Andrich (1978) proposed that each pair of adjacent categories within the rating scale be treated as a series of local dichotomies comprising the categories "more difficult" and "less difficult" to endorse, thereby extending the Rasch model (Rasch, 1980) conceptually. This study examined fit statistics, reliability and separation indices, item-person map, hierarchical ranking of the items, principal components analysis of residuals (PCAR), and the effectiveness of 6-point rating scale categories.

## Results and Discussion

### Fit Statistics

Fit statistics are a quality-control mechanism in Rasch measurement to evaluate two facets of person and item. Two types of fit (“infit” and “outfit”) are calibrated with unstandardized (mean square [MNSQ]) and standardized (ZSTD) values for both facets. As these statistics provide “summaries of Rasch residuals, responses that differ from what is predicted by the Rasch model” (Sick, 2010, p. 24), they are central to determining the unidimensionality of a construct (Bond & Fox, 2007). Although the outfit statistics are unweighted and sensitive to the influence of outlying observations, the infit statistics are weighted and sensitive to patterns in the target responses (Bond & Fox, 2007; Smith, 2001, 2002). As such, Rasch users generally attend to infit rather than outfit values (Bond & Fox, 2007).

Interpretation of fit statistics varies among researchers, and there is no decisive rule for acceptable fit statistics. Although values greater than 1.3 are generally considered misfits (Bond & Fox, 2007), more generous criteria are also used for rating scale (Likert/survey) data. For example, although Wright and Linacre (1994) suggested fit statistics between 0.6 and 1.4 as a reasonable range for rating scale (Likert/survey) data, Linacre (2012) proposed values between .50 and 1.50 to be productive for measurement. This study used the relatively generous criteria proposed by Linacre (2012).

Table 2 summarizes the Rasch item fit statistics of the 12 items that were used to measure the two constructs. All the infit and outfit MNSQ statistics were within the acceptable range (Max. = 1.39, Min. = 0.73 for infit MNSQ statistics; Max. = 1.40, Min. = 0.69 for outfit MNSQ statistics) based on Linacre’s (2012) criteria. Taken together, each item functioned as intended and contributed to measuring the intended unidimensional construct.

Regarding the Rasch person fit statistics, 31 and 41 people were identified to be misfitting with infit MNSQ statistics greater than 1.5 for Group Cohesion and Group Engagement, respectively. A closer look at unexpected responses in the most-misfitting person-response strings revealed no serious problems in their response patterns. For example, the Rasch analysis identified COHE 3 and COHE 1 as unexpected responses for Student 76 with the largest outfit MNSQ statistics (5.92) for Group Cohesion. Student 76 disagreed with COHE 3 (*I enjoy being in the group very much*) but strongly agreed with COHE 1 (*The group is full of laughter*). Because an objective assessment of group work environment is not necessarily comparable to a subjective emotion felt in the group, the response pattern is interpretable. On the other hand, COHE 4 and COHE 6 were identified as unexpected responses for Student 120 with

the second largest outfit MNSQ statistics (4.86) for Group Cohesion. Student 120 agreed with COHE 4 (*I look forward to seeing the group members*) but strongly disagreed with COHE 6 (*Members of the group are personal friends outside the English class*). As group members are not necessarily personal friends beyond the class, this response pattern is also deemed logical.

**Table 2**

*Rasch Item Fit Statistics for the Scale Measuring Group Work Environment*

Item	Measure	SE	Infit		Outfit		PMC
			MNSQ	ZSTD	MNSQ	ZSTD	
<b>Group Cohesion (COHE)</b>							
COHE1	-0.15	0.12	1.39	3.40	1.40	3.33	0.78
COHE6	0.90	0.12	1.22	2.10	1.26	2.28	0.81
COHE3	-0.74	0.12	0.90	-0.97	0.97	-0.27	0.82
COHE4	0.84	0.12	0.83	-1.75	0.81	-1.88	0.85
COHE5	-0.44	0.12	0.83	-1.70	0.86	-1.33	0.83
COHE2	-0.41	0.12	0.73	-2.80	0.69	-3.08	0.86
<b>Group Engagement (ENGA)</b>							
ENGA6	-1.32	0.11	1.25	2.36	1.33	2.89	0.65
ENGA5	0.26	0.10	0.98	-0.12	0.99	-0.03	0.75
ENGA3	0.86	0.10	0.97	-0.29	0.99	-0.08	0.79
ENGA1	0.15	0.10	0.97	-0.29	0.95	-0.50	0.76
ENGA2	-0.18	0.10	0.96	-0.36	0.94	-0.60	0.76
ENGA4	0.24	0.10	0.83	-1.75	0.85	-1.47	0.77

Note. PMC = point-measure correlation.

Concerning Group Engagement, ENGA 6 was identified as an unexpected response for Student 141 with the largest outfit MNSQ statistics (9.72) for Group Engagement. Student 141 disagreed with ENGA 6 (*Members of the group work hard on the class activities*) but agreed with the other five items. This response pattern is puzzling, and he or she might have unintentionally chosen the response or been insincere in his or her responses. Alternatively, three items (ENGA 3, 4, and 5) were identified as unexpected responses for

Student 136 with the second largest outfit MNSQ statistics (5.90). Student 136 agreed with four items (ENGA 1, 2, 3 and 6) but disagreed with two items (ENGA 4 and 5). As shown in the Appendix, whereas the former four items concerned group members' individual engagement on their own tasks, the latter two items tapped group members' concerns about the other members' involvement as well as the progress of the group project. As some groups' members are eagerly engaged in individual activities but do not pay attention to the other members' engagement, the response pattern is not uninterpretable. Taken together, although there were many misfitting students, their response patterns seemed to be logical, to some extent.

### Reliability and Separation Index

Using Rasch analysis, two types of reliability indices were computed for both person and item: reliability and separation. While Rasch person reliability is analogous to Cronbach's  $\alpha$  and indicates the reproducibility of person-measure-order, Rasch item reliability represents the reproducibility of item-measure-order (Linacre, 2012). If the reliability of persons (or items) is high, then "there is a high probability that persons (or items) estimated with high measures actually do have higher measures than persons (or items) estimated with low measures" (Linacre, 2012, p. 575).

The separation index indicates "the number of statistically different levels of performance" (Linacre, 2012, p. 524) in person and item (i.e., person and item separation index). If the person separation index is two, then the instrument is sensitive enough to distinguish the sample on at least two levels. That is, those who exhibit higher endorsement to a given construct are statistically different from those who exhibit the opposite pattern. If the item separation index is three, then three distinct levels are present in terms of item endorsability (easy, average, and difficult).

Table 2 shows the reliability and separation indices for the two constructs. The results for item reliability (separation) were .96 (4.97) and .97 (6.13) for Group Cohesion and Group Engagement, which indicated that approximately a five to six item endorsability hierarchy was present within each construct of the instrument. On the other hand, the results for person reliability (separation) were .88 (2.71) and .83 (2.24) for Group Cohesion and Group Engagement, respectively, which indicated that the instrument was able to distinguish more than two levels for each construct in the sample. It should be noted that the Rasch person reliability values (.88 and .83) were slightly lower than the Cronbach's  $\alpha$  values (.91 and .85) (see the Method section). Although Rasch person reliability is analogous to Cronbach's  $\alpha$ , raw-

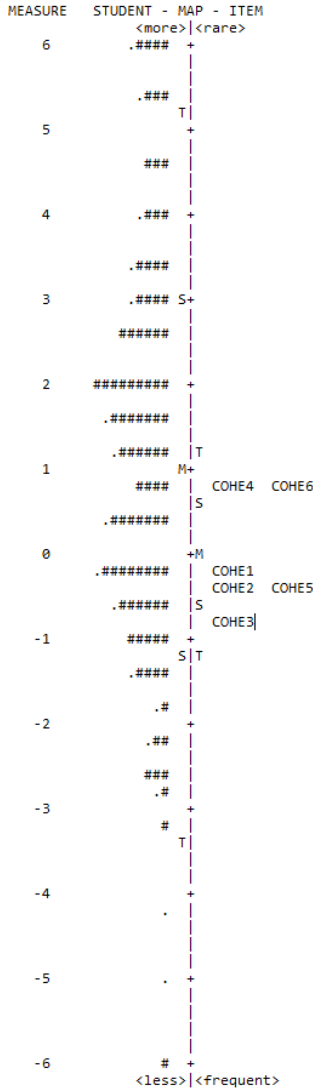
score-based Cronbach's values tend to overestimate reliability, while Rasch underestimates it (Linacre, 2012). Nonetheless, a Rasch person reliability of .88 and .83 is still adequately high. Taken together, the two constructs in the instrument have sufficient reproducibility with adequate power to separate respondents and items into statistically distinguished levels.

### Item-Person Map and Hierarchical Ranking of the Items

Figures 1 and 2 show the Rasch item-person map for Group Cohesion and Group Engagement, respectively. On the far left is the logit scale. The overall distribution of persons is displayed on the left side of the vertical ruler. Respondents who exhibit higher endorsement to a given construct are located higher on the scale. Items' locations are depicted on the right side of the vertical ruler and placed from top to bottom according to the difficulty level of their endorsability; items that are more difficult to endorse are located higher on the map. If a person and an item are located at the same place on the ruler, that person has a 50% probability of endorsing the item on a future iteration of the same measurement instrument. The marker "M" located along the ruler in the maps represents the mean estimate (Linacre, 2012). In general, while the mean item estimate is set to zero, the mean person estimate is relatively calibrated in reference to the items (Sick, 2008b). The instrument's function can be captured by examining the element distribution in both facets (person and item).

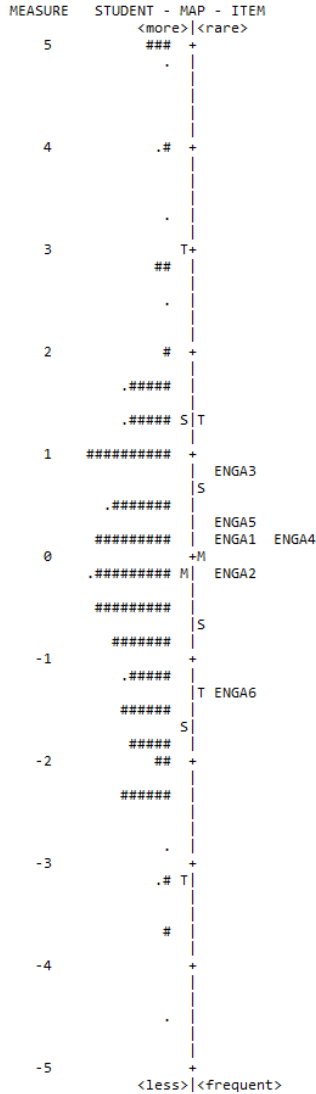
With regard to the construct of Group Cohesion (Figure 1), the mean person ability estimate ( $M = 1.13$ ) was located higher than the mean item endorsability estimate ( $M = .00$ ), which indicated that for some of the participants the items were relatively easy to endorse. Furthermore, there were no equivalent levels of items to distinguish respondents located around the upper part of the map (i.e., respondents with a logit scale score above 1.00). Ideally, questionnaire items should cover "a range of difficulty that matches the range of person measures in the target audience" (Sick, 2011, p. 16). Although the instrument was highly reliable with a reliability estimate of .96 and .88 for item and person, respectively, it lacked the power to separate the sample into three distinguished levels with a person separation index of less than three. Inclusion of more difficult items to endorse would differentiate the respondents with higher measures and increase the accuracy of the assessment. Furthermore, as some items overlapped on the item hierarchy (i.e., COHE 4 and 6, and COHE 2 and 5), one of the two items exhibiting similar endorsability may be replaced with a more difficult item to endorse.

**Figure 1**  
*Item-Person Map Depicting Respondents' Agreeability and Item Endorsability for the Construct of Group Cohesion*



*Note.* The number sign (#) and dot (.) represent two respondents and one respondent, respectively. M = mean; S = one standard deviation away from the mean; T = two standard deviations away from the mean.

**Figure 2**  
*Item-Person Map Depicting Respondents' Agreeability and Item Endorsability for the Construct of Group Engagement*



*Note.* The number sign (#) and dot (.) represent two respondents and one respondent, respectively. M = mean; S = one standard deviation away from the mean; T = two standard deviations away from the mean.



A closer look at the item endorsability revealed that the most difficult item to endorse was COHE 6 with 0.90 logits (*Members of the group are personal friends outside the English class*), followed by COHE 4 with 0.84 logits (*I look forward to seeing the group members*), COHE 1 with -0.15 logits (*The group is full of laughter*), COHE 2 with -0.41 logits (*Members of the group get along with each other*), COHE5 with -0.44 logits (*I like the group*), and COHE 3 with -0.74 logits (*I enjoy being in the group very much*). As the difference between COHE 6 and 4 was only 0.06 logits, endorsability of these two items is virtually the same. Regarding COHE 6, the students stayed in the same group over a semester. Although some members formed friendships beyond the English class, others did not. As greater cohesion is required to build a friendship beyond the class, it is reasonable that COHE 6 was ranked as the most difficult item to endorse. Concerning COHE 4, looking forward to seeing the group members requires stronger emotional involvement than merely liking the group (i.e., COHE 5) and objective assessment of cohesive group work environment (i.e., COHE 1 and 2). Hence, it is also considered logical that COHE 4 was the most difficult item to endorse. Given the items' content, the item measure hierarchy of the remaining items was also logical, which verified construct validity.

With regard to the construct of group engagement (Figure 2), the mean person ability estimate ( $M = 0.04$ ) closely matched the mean item difficulty estimate ( $M = 0.00$ ); thus, overall, the items were well targeted for the participants. The most difficult item to endorse was ENGA 3 with 0.86 logits (*Members of the group work on the tasks and activities more than the teacher requires*). Given that greater engagement is necessary to perform a task beyond the teacher's requirement, it is deemed logical that ENGA 3 was the most difficult item to endorse. ENGA 5 (*Members of the group care about whether the other members are doing well on the activities*) and ENGA 4 (*Members of the group have great concern for the progress of group activities*) exhibited similar endorsability (0.26 and 0.24 logits, respectively) and were the second most difficult items to endorse. These items tapped group members' concerns about the other members' involvement, as well as the progress of a group project as a whole. The remaining easier items to endorse, e.g., ENGA 1 (*Members of the group are highly motivated to perform the English project*; 0.15 logits) and ENGA 2 (*Members of the group prepare for and practice the presentations well*; -0.18 logits), focused on group members' individual engagement with their own tasks. As greater involvement is generally required to care about the group as a whole than an individual's own task, the item hierarchy seemed understandable. The easiest item to

endorse was ENGA 6 with  $-1.32$  logits (*Members of the group work hard on the class activities*). Whereas the other items require greater engagement entailing out-of-class activities (e.g., preparation and practice of a presentation), this item limited the scope to only in-class activities. This may be why ENGA 6 was a very easy item to endorse, with a wide gap (1.14 logits) to the second easiest item (i.e., ENGA 2). Taken together, the item measure hierarchy was deemed logical for group engagement, thus verifying construct validity.

### Rasch PCAR

As with conventional factor analysis, the Rasch PCAR is used to analyze the unidimensionality of the construct. The purpose of factor analysis is to “summarize patterns of correlations among observed variables [and] reduce a large number of observed variables to a smaller number of factors” (Tabachnick & Fidell, 2007, p. 608), that is, to construct factors. On the other hand, the purpose of the Rasch PCAR is not to find shared factors but to examine whether residuals after the removal of the primary measurement dimension share a substantive attribute in common to form a secondary dimension. Unidimensionality of construct can be confirmed by “a failure to find any meaningful components beyond the primary dimension of measurement” (Sick, 2010, p. 24) in the Rasch analysis. According to Linacre (2012), as “the variance of two items” is required to form a secondary dimension, a construct is unidimensional if the first residual contrast has an eigenvalue of less than 2.0. Furthermore, the first residual contrast should account for less than 10%. The results of the analysis showed that the Rasch model explained more than half the variance for both constructs (65.7% and 56.8% for Group Cohesion and Group Engagement, respectively). The first residual contrast (eigenvalue) accounted for 9.0% (1.6) and 11.3% (1.6) for Group Cohesion and Group Engagement, respectively. Although the eigenvalues were well below the criterion of 2, the value of 11.3% for Group Engagement was slightly larger than the criterion. In order to examine a possible cause, 9 highly misfitting people, with infit MNSQ values larger than 2.5 (i.e., less than 5% of 200 persons), were temporarily eliminated and recalibrated. Results revealed that the first residual contrast (eigenvalue) only explained 9.2% (1.4). Given that less than 5% of the students with large misfit values caused the inflation of the first residual contrast, this is not considered a problem. Taken together, these findings demonstrated the presence of a reasonable amount of the primary dimension and a lack of the secondary dimension for both constructs, thus confirming the unidimensionality of both constructs.

### Rating Scale Categories

The participants of this study responded using 6-point rating scale categories (1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Slightly disagree*, 4 = *Slightly agree*, 5 = *Agree*, and 6 = *Strongly agree*). The effectiveness of the rating scale categories was assessed based on guidelines proposed by Linacre (2002) and Wolfe and Smith (2007). Results revealed problems in Category 1 (i.e., Strongly disagree) and Categories 1 and 2 (i.e., Strongly disagree and Disagree) in the original six-point rating scales for Group Cohesion and Group Engagement. Combining the two scale categories (i.e., 1 = *Strongly disagree*, and 2 = *Disagree*) led to an improved quality of the rating scales, suggesting that the use of five-point scales was optimal for the respondents. Please note that all the analyses in the preceding sections were conducted using the optimal five-point rating scale.

### Conclusion

This study aimed to evaluate the construct validity of the scale used to assess a group work environment (Tanaka, 2017) using the Rasch model. Although the quality of the scale has room for improvement (e.g., by adding more difficult endorsement items for the group cohesion construct), the results of the Rasch analysis verified construct validity with acceptable fit statistics, adequately high reliability and separation indices, logical hierarchical ranking of the items, and a demonstration of the unidimensionality of the construct. Thus, in answer to the research question posed in this study, the questionnaire instrument of Tanaka (2017) was demonstrated to indeed measure what it was intended to. As shown in prior research (Poupore, 2016, 2018; Tanaka, 2018), the group work environment plays an influential role in the foreign language classroom. Although, according to the content of the group work, some item revision may be necessary, use of the questionnaire (Tanaka, 2017) can help teachers gauge the function of each group work environment. It also provides researchers with a tool to investigate the role of the group work environment in relation to various L2 learning aspects, including individual difference variables such as motivation.

**Mitsuko Tanaka** is an associate professor at Osaka University. Her current research interests include individual differences in SLA (e.g., motivation and self-construal) and language assessment.

## References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573. <https://doi.org/10.1007/BF02293814>
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Erlbaum.
- Chang, L. Y. H. (2010). Group processes and EFL learners' motivation: A study of group dynamics in EFL classrooms. *TESOL Quarterly*, 44(1), 129-154. <https://doi.org/10.5054/tq.2010.213780>
- Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, 44(3), 417-448. <https://doi.org/10.1111/j.1467-1770.1994.tb01113.x>
- Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *The Modern Language Journal*, 78(3), 273-284. <https://doi.org/10.2307/330107>
- Dörnyei, Z., Henry, A., & Muir, C. (2016). *Motivational currents in language learning: Frameworks for focused interventions*. Routledge.
- Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4(3), 275-300. <https://doi.org/10.1177/13621688000400305>
- Dörnyei, Z., & Kubanyiova, M. (2014) *Motivating learners, motivating teachers*. Cambridge University Press.
- Dörnyei, Z., & Murphey, T. (2003). *Group dynamics in the language classroom*. Cambridge University Press.
- Dörnyei, Z., & Ushioda, E. (2013). *Teaching and researching motivation* (2nd ed.). Routledge.
- Ito, A., & Matsui, H. (2001). Construction of the classroom climate inventory. *Japanese Journal of Educational Psychology*, 49(4), 449-457. [https://doi.org/10.5926/jjep1953.49.4\\_449](https://doi.org/10.5926/jjep1953.49.4_449)
- Joe, H-K., Hiver, P., & Al-Hoorie, A. H. (2017). Classroom social climate, self-determined motivation, willingness to communicate, and achievement: A study of structural relationships in instructed second language settings. *Learning and Individual Differences*, 53, 133-144. <https://doi.org/10.1016/j.lindif.2016.11.005>
- Kozaki, Y., & Ross, S. J. (2011). Contextual dynamics in foreign language learning motivation. *Language Learning*, 61(4), 1328-1354. <https://doi.org/10.1111/j.1467-9922.2011.00638.x>

- Leeming, P. (2020) Investigating collective-efficacy in the foreign language classroom. *The Language Learning Journal*, 48(2), 237-252. <https://doi.org/10.1080/09571736.2017.1416424>
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2012). *A user's guide to WINSTEPS: Rasch-model computer program*. Winsteps.com.
- Linacre, J. M. (2013). Winsteps (Version 3.80.0) [Computer software]. Winsteps.com.
- Poupore, G. (2016). Measuring group work dynamics and its relation with L2 learners' task motivation and language production. *Language Teaching Research*, 20(6), 719-740. <https://doi.org/10.1177/1362168815606162>
- Poupore, G. (2018). A complex systems investigation of group work dynamics in L2 interactive tasks. *The Modern Language Journal*, 102(2), 350-370. <https://doi.org/10.1111/modl.12467>
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. (Expanded ed.) University of Chicago Press. (Original work published 1960)
- Sasaki, M., Kozaki, Y., & Ross, S. J. (2017). The impact of normative environments on learner motivation and L2 reading ability growth. *The Modern Language Journal*, 101(1), 163-178. <https://doi.org/10.1111/modl.12381>
- Sick, J. (2008a). Rasch measurement in language education, Part 1. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 12(1), 1-6.
- Sick, J. (2008b). Rasch measurement in language education, Part 2: Measurement scales and invariance. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 12(2), 26-31.
- Sick, J. (2010). Rasch measurement in language education, Part 5: Assumptions and requirements of Rasch measurement. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 14(2), 23-29.
- Sick, J. (2011). Rasch measurement in language education, Part 6: Rasch measurement and factor analysis. *Shiken: JALT Testing and Evaluation SIG Newsletter*, 15(1), 15-17.
- Smith, E. V. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. *Journal of Applied Measurement*, 3(2), 205-231.

- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson Education.
- Tanaka, M. (2017, November 17-20). Measuring Group Work Dynamics [Paper presentation]. The 43rd Annual International Conference of the Japan Association for Language Teaching (JALT), Tsukuba International Congress Center, Tsukuba, Japan.
- Tanaka, M. (2018, June 7-10). Individual perceptions of group work environment and L2 learning motivation [Paper presentation]. The 3rd International Psychology of Language Learning Conference (PLL3), Waseda University, Tokyo, Japan.
- Wolfe, E. W., & Smith, E. V. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—validation activities. *Journal of Applied Measurement*, 8(2), 204-234.
- Wright, B. D., & Linacre, M. (1994). Reasonable mean square fit values. *Rasch Measurement Transactions*, 8(3), 370. <http://www.rasch.org/rmt/rmt83b.htm>

## Appendix

### English Translations of Questionnaire Items about the Group Work Environment (Tanaka, 2017)

To what extent do you agree with each of the following statements about your group?

---

#### ***Group Cohesion (COHE)***

---

- |       |  |
|-------|--|
| COHE1 | The group is full of laughter.                                       |
| COHE2 | Members of the group get along with each other.                      |
| COHE3 | I enjoy being in the group very much.                                |
| COHE4 | I look forward to seeing the group members.                          |
| COHE5 | I like the group.  |
| COHE6 | Members of the group are personal friends outside the English class. |
- 

#### ***Group Engagement (ENGA)***

---

- |       |   |
|-------|---|
| ENGA1 | Members of the group are highly motivated to perform the English project.                   |
| ENGA2 | Members of the group prepare for and practice the presentations well.                       |
| ENGA3 | Members of the group work on the tasks and activities more than the teacher requires.       |
| ENGA4 | Members of the group have great concern for the progress of group activities.               |
| ENGA5 | Members of the group care about whether the other members are doing well on the activities. |
| ENGA6 | Members of the group work hard on the class activities.                                     |
- 

*Note.* All the questionnaire items are randomly ordered 6-point Likert scale items.