

科学研究費助成事業 研究成果報告書

平成 29 年 9 月 3 日現在

機関番号：32408

研究種目：基盤研究(C) (一般)

研究期間：2013～2016

課題番号：25370726

研究課題名(和文) 教員採用試験における模擬授業評価者の評価過程分析

研究課題名(英文) An investigation of rater variability in the context of teacher employment examinations using a mixed approach

研究代表者

秋山 朝康 (Akiyama, Tomoyasu)

文教大学・文学部・教授

研究者番号：20383218

交付決定額(研究期間全体)：(直接経費) 2,100,000円

研究成果の概要(和文)：公立の教員採用試験の受験者は全国で約18万人で、最終倍率は各教科や都道府県市で異なるが英語科(中・高)の平均採用率は10%台である。本研究は評価者の評価の過程を調べるために評価者17人に模擬授業を採点してもらい、量的分析(ラッシュ分析)と質的分析(シンク・アウド法)を行った。量的分析では多くの評価者は一貫して評価しているがそれぞれの「厳しさ」で評価しているため実際の教員採用試験で、誰が評価者になるかにより合否が決まってしまうことは否定できない。質的分析結果は評価項目の使用度合いや評価項目の解釈は評価者によって多種多様であり、そのことが評価者間の違いや評価のぶれに繋がっていると考えられる。

研究成果の概要(英文)：This study aims to examine possible reasons of rater variability in employing prospective English teachers, both quantitatively (Many-Facet Rasch Analysis, MFRA) and qualitatively (think-aloud method). Thirty candidates undertook microteaching tests. Seventeen raters rated all candidates using assessment criteria. Results indicate that most of the raters show consistent rating patterns but with various levels of severity. Results of the think-aloud method show that raters rated candidates not only with different interpretations of and different attention to subcategories of assessment criteria but also according to different core teaching values and ideal images of English teachers or English classes possessed by the rater themselves, leading to divergence between lenient and harsh ratings of the same candidates by different raters.

研究分野：言語テスト

キーワード：教員採用試験 パフォーマンステスト 模擬授業 評価者

1. 研究開始当初の背景

(1) 公立の教員採用試験の受験者は全国で約 18 万人である。最終倍率は各教科や都道府県市で異なるが中・高英語科の平均採用率は 10% 台である (文科省、2016)。このような重要なテストにも関わらず教員採用試験に関する研究は数えるほどしかない。

(2) 教員採用試験の 2 次試験では主に面接 (個人・集団) や小論文や模擬授業などが実施されていて、2 次試験の結果は直接、採用の可否に関わる。Akiyama (2011) は 2 次試験で実施される「模擬授業」についてほとんど研究されていないことを指摘し、教員採用試験で模擬授業を使用する課題 (例えば、評価者の信頼性、短時間で受験者の授業力を評価することの妥当性) を指摘した。

2. 研究の目的

本研究は模擬授業の評価者に焦点をあて、主に以下の研究課題を設定した。

- (1) 模擬授業評価者間にどのような差異があるのか?
- (2) 評価者はどのように評価項目を用いているのか (なぜ評価 (採点) に差異が生じるのか?)

3. 研究の方法

(1) 研究参加者は 30 人の大学生の模擬授業者と 17 人の評価者である。評価項目は 6 つ (1. 授業の構成力・2. 指導力・3. 表現力・4. 教師の資質・5. 教科の専門性・6. 採用するか否かの全体的な評価) である。

・模擬授業者は事前に与えられた文法導入項目について 20 分間考えた。紙とマジックは自由に使えるようにした。授業は 5 分間を超えることはないが 5 分以内に終わった場合は終了とした。

・評価者 17 人は以下の 3 つのグループに分けて評価してもらった。(J グループは中学英語教師 6 人、M グループは指導主事経験者・管理職 6 人、S グループは高校英語教師 5 人) 評価の採点は各項目 1 点~5 点で、最低点は

6 点、最高点は 30 点である。

(2) 分析方法

研究課題を分析するために、2 種類のデータを分析した。一つは評価者のスコアを分析するために Rasch analysis (FACETS) を用いた。評価者の評価の一貫性、厳しさ、バイアスなどを分析した。もう一つは評価者が評価しているときになぜその点数を与えたのかを声になだしてもらい IC レコーダーに吹き込んでもらった。(Think-aloud method)。このデータを NVivo を用いて分析した。

4. 研究成果

(1) 評価者間に評価の一貫性と採点の厳しさにどのような違いがあるのか。

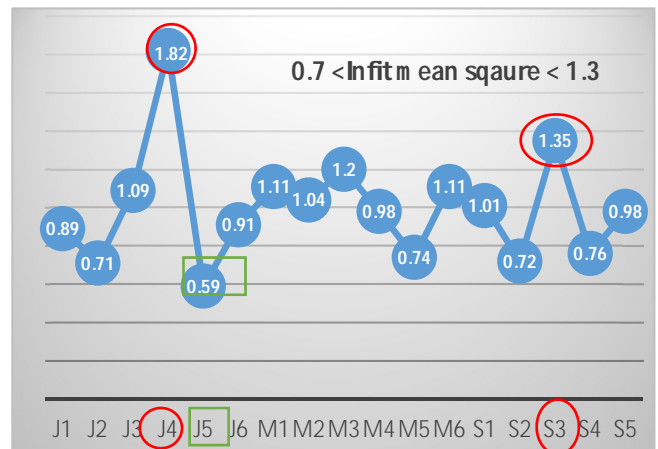


図 1 : 17 人の評価者の評価の一貫性

図 1 はラッシュ分析で評価者 17 人のスコアの信頼性 (一貫性) を示した図である (J は中学英語教師、M は元英語教師で現在は管理職、S は高校英語教師)。この図では数値 (infit mean square) が 0.7 以上 1.3 未満であれば評価者のスコアは予想の範囲であると結論付ける。つまり、評価者の評価は比較的一貫性があり安定しているという指標である。図 1 から 2 人の評価者 (J4 & S3) が上限値を越え、J5 が下限値を下回っている。前者はラッシュが予測したデータパターンが評価者とはかけ離れていることを示唆し、後者は評価者データパターンがラッシュの

予測したデータパターンよりバリエーションが少ないことを示している。残りの14人の評価者は比較的一貫性のある評価をしていると推定できる。

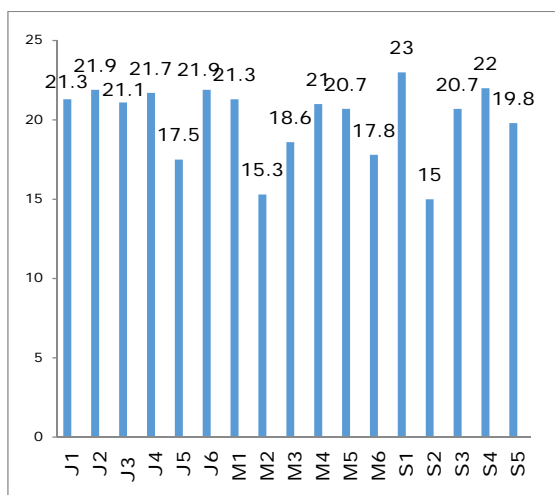


図2 17人の評価者の厳しさ(平均点)の違い

図2は17人の評価者の平均点を表している。低い評価者は厳しく、高い評価者は易いことを示す。最高点が30点で最低点が6点である。もっとも厳しい評価者はS2(15点)やM2(15.3点)であった。易い評価者はS1(23点)S5(22点)で評価者の差の平均は約7点であった。このことは評価者によって違いが大きく採用試験の合否に影響があることを示唆している。つまり図1と図2の結果を合わせると、多くの評価者は一貫して評価しているようであるがそれぞれの「厳しさ」で評価しているため実際の教員採用試験で、誰が評価者になるかにより合否は決まってしまうことは否定できないと言える。

(2) 評価者はどのように評価項目を使用しているか。

図3は17人の評価者がどのように評価項目を使用していたか think-aloud method を用いて分析したものである。各評価者には2本のグラフがあるが、左側が実際に提示された

評価項目、右側が評価項目以外の項目(評価者オリジナルの評価項目の)の使用度合いを示している。

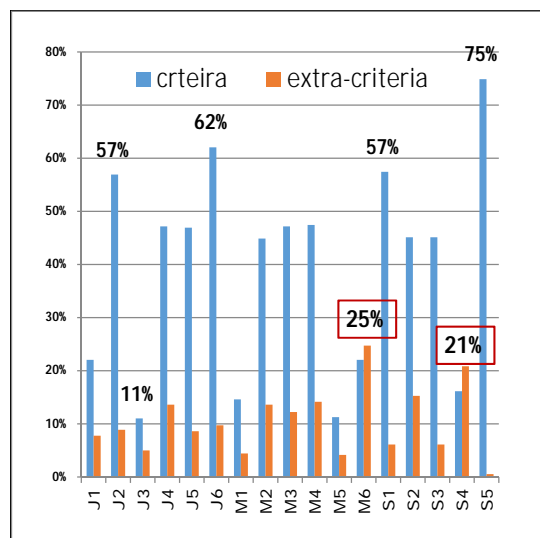


図3 17人の評価項目の使用結果

図3で特徴的なことは3点ある。一つ目はどの評価者も用意された評価項目をすべて使用しているわけではないということ。特に、M1とM5は評価項目の使用度合いが20%以下であった。次に大きく分けると2種類のタイプの評価者が存在すること。用意された評価項目を使用している評価者(J2, J6, S1 & S5)と評価者が作りだしたオリジナル評価の使用者(J3, M6 & S4)がいる。次にS5を除くすべての評価者は評価項目に自分の解釈を加えて評価しているとのこと。これらの結果をまとめると同じ評価項目を使用してもその使用度合いや解釈は評価者によって多種多様であり、評価者間で違う評価になっていることが想像できる。

(3) 評価者はどのように評価項目を使用しているか。

下記の表は受験者No6の模擬授業に対して2人の評価者が評価項目をどれくらい使用しているかを比較したものである。表からJ6はJ4よりもより多く評価項目を使用している。そして3つの項目(LF3, D3, E3,03)は

評価項目以外のことを指しているので J4 は用意された評価項目はあまり使用せず、評価者自身が拡大解釈又は新しく評価項目を付け足して評価項目を使用していることがわかる。また下の結果ではすべての項目で P1 の部分だけが共通に使用していることがわかった。同じ評価項目を用い、同じ授業を見ているでも違う観点で評価していると言える。

A comparison of two raters' criteria use on No.6

J6				J4			
LF		LF2	LF3	LF	LF1		
IA	IA1		IA3	IA		IA2	
D	D1	D2		D			D3
E	E1			E			E3
P	P1	P2		P	P1		
O		O2		O			O3

また上記の結果は図 1 で示されたように J4 は最も信頼性が低い評価者であった。上の表の結果は J4 が評価項目を使用していないことと関係するかもしれない。

しかしながら評価者が声にださないことを評価していないと考えるべきではない。そこで think-aloud 法の弱点を補うためにアイトラッキング技術を応用し目の動きを分析しどのように評価しているかを調査することが必要である。しかしアイトラッキング技術もなぜそのスコアにしたのかを明らかにすることは難しい。そこで上記の二つの方法をお互いに補完するために、今後はアイトラッキング技術と think-aloud 法、両方用いて、さらに採点後のインタビューやアンケートを用いて評価過程を調べ、少しでも公平な評価が実施されるよう研究を続けていくことが必要である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 件)

- ① Akiyama, T (2013). An investigation of test fairness to prospective English teachers in Teacher Employment Examinations (TEEs): focusing on bias interactions between raters and candidates, Annual Review of English Language Education Journal, 24, pp279-294.
- ② Murry, A. Akiyama, T. at al. (2013). Fundamental Issues Surrounding Integrated Tests in Terms of Assessment Literacy- The Case of Integrated Speaking Tests -, PAAL proceedings, pp1-11.
- ③ 秋山朝康 (2014). 採用試験の評価者は模擬授業をどのように評価しているのか -think-aloud 法を用いて-「英語英文学」41号, pp 9-22.
- ④ Akiyama, T. (2014) Assessing Speaking: Issues and problems with Rater Cognition Research, JACET Testing SIG Book No.2: Speaking Monograph, pp19-34.
- ⑤ Akiyama, T. (2016). Assessing teacher, 20th Special Issue of JLTA Journal. Pp168-172.

〔学会発表〕(計 10 件)

- ① Akiyama, T. (2013). Issues with integrative teaching & assessment in Japan: From assessment (exit) to teaching (entrance) Invited Panelist, COLT 2nd International Conference on Language Learning and Teaching, Paper presented at the CoLT 2013, Park Royal Hotel Batu Feringghi, Penang, Malaysia.
- ② Akiyama, T. (2013). Investigations of Teacher Employment Examinations: How do Raters Assess Prospective English teachers? Paper presented at the COLT 2nd International Conference on Language Learning and Teaching (CoLT 2013) Park Royal Hotel Batu Feringghi, Penang, Malaysia.
- ③ Matsumoto, K., Akiyama, T., Nakamura, Y., Ito, Y., and Miyazaki, K. (2013). Issues Surrounding Integrated-skills Assessment in Japan - An Attempt toward the Improvement of Integrated-skills Speaking Tests, Paper presented at the ASIA TEFL, Philippine.
- ④ 秋山朝康(2013) 評価者はどのように評価

しているのか～教員採用試験の模擬授業の場合
全国英語教育学会（北星学園）

⑤ Akiyama, T. (2013). "Where Raters' Differences Come from in the Context of English Teacher Employment Examinations via microteaching". Paper presented at the American Association for Applied Linguistics 2013 Conference in Dallas, Texas (March 16-19).

⑥ Akiyama, T. (2014). The possibility of introducing speaking tests into senior high school entrance examinations, Enhancing the validity and practicality of speaking assessment. *Invited Symposium, JLTA conference*, October, Ritsumeikan University.

⑦ Nakamura, Y. Akiyama, T., Ito, Y., Matsumoto, K., Miyazaki, K. Murray, A. and Tsuchihira, T. (2014), The Establishment of Models for Integrated-skills Assessment. From the Standpoint of Assessment Literacy, Paper presented at the AILA conference, August, Brisbane, Australia.

⑧ Akiyama, T. (2014) "Who Assesses Tomorrow's English Teachers?" An impact of Rater Differences on Teacher Employment Decisions, Paper presented at the AILA conference, August, Brisbane, Australia.

⑨ Akiyama, T. (2014). Examining test fairness: Focusing on Rater Behavior in the Case of Employing Prospective English Teachers, Paper ALTE The Association of Language Testers in Europe at the Maison Internationale, Paris, France

⑩ Akiyama, T. (2015). Investigations of teacher employment examinations: How do raters assess prospective English teachers? Applied Linguistics Association of Australia (ALAA), the Applied Linguistics Association of New Zealand (ALANZ) and the Association for Language Testing and Assessment of Australia and New Zealand (ALTAANZ), Paper will be presented, November 30th, University of South Australia,

AUS.

11 Akiyama, T. (2015). An exploration of rater cognition: The impact of rater differences on teacher employment examinations. October 9th. Paper presented at the 2015 ECOT conference, Georgetown University, USA.

12 Akiyama, T. (2015). An Examination of Rater Orientations in English Teacher Employment Examinations, Paper will be presented at the CAES International Conference, in University of Hong Kong, HK.

13 Akiyama, T. (2016). An Investigation of Rater Variability in the Context of Teacher Employment Examinations Using a Mixed Approach. Paper presented at International Symposium on Language, Linguistics, Literature and Education., Okinawa, Japan.

14 Akiyama, T. (2016). A Closer Take a Look at Rater Bias in English Teacher Employment Examinations. Paper presented at PAAL (The Pan-Pacific Association of Applied Linguistics), Tamsui Campus of Tamkang University, Taiwan.

15 Akiyama, T. (2016). "Searching into rater variability in employing English teachers using multiple analytic approaches". Paper presented at the American Association for Applied Linguistics 2017 Conference in Portland, Oregon, USA (March 18-21).

〔図書〕(計 1 件)

① Akiyama, T. (2016). Test fairness to prospective English teachers in Teacher Employment Examinations (TEEs): An investigation focusing on rater behavior in Studies in Language Testing vol. 44, *Language Assessment for Multilingualism* pp. 372-396, Cambridge University Press.

6. 研究組織

(1) 研究代表者

秋山朝康 (AKIYAMA Tomoyasu)

文教大学・文学部・教授

研究者番号 20383218