L2 Oral Fluency
Towards Bridging a Gap Between SLA Research and Language Testing

Shungo Suzuki, Ph.D.
Assistant research professor, Waseda University
Email: ssuzuki@pcl.cs.waseda.ac.jp
Twitter: @shungosuzuki

11:40-12:25 (GMT +09:00), 27th of November, 2021
The Japan Association for Language Education and Technology, Methodology SIG
My current project

Research Topic 1
Massive Conversational Data Collection and Analysis Ecosystem

Research Topic 2
Dialog-based Automated Speaking Skill Assessment

Research Topic 3
Massive Conversational Data Collection and Analysis Ecosystem
My current project

An automated-scoring system of 6 dimensions of speaking performance (incl. fluency) based on the CEFR

HOWEVER...

I’ve faced research gaps in L2 fluency research.
Fluency as an important construct in testing

**IELTS**

**TOEFL iBT® Test**

**Independent Speaking**

<table>
<thead>
<tr>
<th>Band</th>
<th>Score</th>
<th>General Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>5</td>
<td>Native-like</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

**Oral fluency**

- **5 Native-like**
  - Speech shows smooth rhythm and phrasing. There are no hesitations, repetitions, false starts or non-native phonological simplifications.
- **4 Advanced**
  - Speech has an acceptable rhythm with appropriate phrasing and word emphasis. There is no more than one hesitation, one repetition or a false start. There are no significant non-native phonological simplifications.
- **3 Good**
  - Speech is at an acceptable speed but may be uneven. There may be more than one hesitation, but most words are spoken in continuous phrases. There are few repetitions or false starts. There are no long pauses and speech does not sound staccato.
- **2 Intermediate**
  - Speech may be uneven or staccato. Speech (if >= 6 words) has at least one smooth three-word run, and no more than two or three hesitations, repetitions or false starts. There may be one long pause, but not two or more.
How have SLAers conceptualized fluency?
Tavakoli & Hunter (2018)

- Focus: Teachers’ understanding of fluency
- Who?: L2 teachers in the UK (N = 84)
- Data: Questionnaire (QUAN + QUAL)
- Finding:
  - Temporal fluency...13.4%
  - Speaking ability...43.8%
- SLA research: A very narrow perspective
- Suggestion: 4 different perspectives

“As regards language benchmark documents such as the CEFR..., although, in principle, they encourage a narrower perspective to fluency..., they may not provide teachers with an adequately clear and easy-to-work-with framework for defining fluency and promoting it in the classroom.” (p. 342)
Three-dimensional model of fluency

- Tavakoli & Skehan (2005): Only based on picture narrative speech
- Suzuki & Kormos (in press): Confirmed its generalizability across four different tasks

How useful are they in language testing?
Suzuki, Kormos, & Uchihara (2021)

• Focus: Association between subjective fluency ratings and temporal features

• Method: Meta-analysis based on 263 effect sizes (22 studies)
  • N.B. Only those based on monologic speech were included

• 6 selected measures
  • Articulation rate (Speed)
  • Pause frequency & duration (Breakdown)
  • Disfluency rate (Repair)
  • Speech rate & Mean length of run (Composite)

The aggregated correlation coefficients

Perceived fluency

1. Composite measures: $r = 0.72 - 0.76$
2. Speed fluency: $r = 0.62$
3. Breakdown fluency (frequency): $r = 0.59$
4. Breakdown fluency (duration): $r = 0.46$
5. Repair fluency: $r = 0.20$
The aggregated correlation coefficients

Perceived fluency

- Mid-clause pause measures: $r = .71-.72$
- Composite measures: $r = .72-.76$
- Speed fluency: $r = .62$
- Breakdown fluency (frequency): $r = .59$
- Breakdown fluency (duration): $r = .46$
- Repair fluency: $r = .20$
R² values with multiple regression

- Min: 50.9%
- Median: 66.5%
- Max: 94.0%
What we’ve seen in SLA research

• SLA research has adopted a very narrow perspective, which consists of three sub-dimensions:
  • Speed fluency
  • Breakdown fluency
  • Repair fluency

• Existing fluency measures can explain the 51–94% of variance of subjective ratings of fluency in monologic speech.

• Little is known about the predictive power of temporal features in fluency ratings in dialogic speaking tasks (see Tavakoli & Wright, 2020).

How have Language Tests assessed fluency?
Some high-stakes tests...

<table>
<thead>
<tr>
<th>Tests</th>
<th>Named Construct</th>
<th>Highest levels</th>
<th>Speaking tasks</th>
</tr>
</thead>
</table>
| IELTS          | Fluency and Coherence            | • speaks fluently with only rare repetition or self-correction; any hesitation is content-related rather than to find words or grammar  
• speaks coherently with fully appropriate cohesive features  
• develops topics fully and appropriately                     | Interview            |
| ETS TOEFL      | Delivery                         | Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility. | Monologue            |
| PTE Academic   | Oral fluency                     | Speech shows smooth rhythm and phrasing. There are no hesitations, repetitions, false starts or non-native phonological simplifications.         | Short monologue      |
Some high-stakes tests...

<table>
<thead>
<tr>
<th>very narrow (speed, breakdown and repair)</th>
<th>Speaking tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>narrow (ease, flow and continuity; distinct from accuracy and complexity)</td>
<td>Interview</td>
</tr>
<tr>
<td><strong>broad</strong> (L2 speaking ability)</td>
<td><strong>Fluency + Coherence</strong></td>
</tr>
<tr>
<td><strong>very broad</strong> (mastery; general proficiency)</td>
<td><strong>Fluency + Pronunciation</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fluency + Coherence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fluency + Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monologue</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Short monologue</th>
</tr>
</thead>
</table>

---

Fluency + Coherence

Fluency + Pronunciation

Somewhere in-between “narrow” and “broad” perspectives to fluency?
Fluency scale in CEFR (2018, p.146)

“... the scale below focuses more on the narrower ( = accessing one’s repertoire), more traditional view of fluency. ...

Key concepts operationalised in the scale include the following:
• Ability to construct utterances, despite hesitations and pauses (lower levels);
• Ability to maintain a lengthy production or conversation;
• Ease and spontaneity of expression.”

• C2 Can express themselves at length with a natural, effortless, unhesitating flow. Pauses only to reflect on precisely the right means to express their thoughts or to find an appropriate example or explanation.

• C1 Can express themselves fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.

• B2 Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of language. Can produce stretches of language with a fairly even tempo; although they can be hesitant as they search for patterns and expressions, there are few noticeably long pauses. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party.

• B1 Can express themselves with relative ease. Despite some problems with formulation resulting in pauses and “cul-de-sacs”, they are able to keep going effectively without help. Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.

• A2 Can make themselves understood in short contributions, even though pauses, false starts and reformulation are very evident. Can construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts.

• A1 Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words/signs, and to repair communication.
• C2 Can express themselves at length with a natural, effortless, unhesitating flow. Pauses only to reflect on precisely the right means to express their thoughts or to find an appropriate example or explanation.

• C1 Can express themselves fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.

• B2 Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of language. Can produce stretches of language with a fairly even tempo; although they can be hesitant as they search for patterns and expressions, there are few noticeable long pauses. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party.

• B1 Can express themselves with relative ease. Despite some problems with formulation resulting in pauses and "cul-de-sacs", they are able to keep going effectively without help. Can keep going comprehensively, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.

• A2 Can make themselves understood in short contributions, even though pauses, false starts and reformulation are very evident. Can construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts.

• A1 Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words/signs, and to repair communication.
• C2 Can express themselves at length with a natural, effortless, unhesitating flow. Pauses only to reflect on precisely the right means to express their thoughts or to find an appropriate example or explanation.

• C1 Can express themselves fluently and spontaneously, almost effortlessly. Only a conceptually difficult subject can hinder a natural, smooth flow of language.

• B2 Can communicate spontaneously, often showing remarkable fluency and ease of expression in even longer complex stretches of language. Can produce stretches of language with a fairly even tempo; although they can be hesitant as they search for patterns and expressions, there are few noticeably long pauses. Can interact with a degree of fluency and spontaneity that makes regular interaction with users of the target language quite possible without imposing strain on either party.

• B1 Can express themselves with relative ease. Despite some problems with formulation resulting in pauses and "cul-de-sacs", they are able to keep going effectively without help. Can keep going comprehensibly, even though pausing for grammatical and lexical planning and repair is very evident, especially in longer stretches of free production.

• A2 Can make themselves understood in short contributions, even though pauses, false starts and reformulation are very evident. Can construct phrases on familiar topics with sufficient ease to handle short exchanges, despite very noticeable hesitation and false starts.

• A1 Can manage very short, isolated, mainly pre-packaged utterances, with much pausing to search for expressions, to articulate less familiar words/signs, and to repair communication.

We may also consider...
- Conceptual difficulty/elaboration
- Linguistic complexity (see Skehan, 2009)
- Different weights across levels (see Tavakoli et al., 2020)


Towards automated-scoring of oral fluency
Demo movie: InteLLA

• https://www.youtube.com/watch?v=RzCq5Z4cDBk
NLP, Dialogue system and Machine learning

• NLP: Automated annotation of temporal features
  • Speech recognition
  • Pruning (annotation for disfluencies)
  • Pause detection
  • Annotation for pause locations
  • The software will be released soon! (Matsuura, Suzuki, et al., forthcoming)

• Adaptive dialogue system—InteLLA (Saeki et al., 2021)
  • Ratability:
    • Amount of speech elicited
    • Appropriate difficulty levels
  • Incremental assessment + Adaptive interview strategies

NLP, Dialogue system and Machine learning

• Neural Network
  • Optimizing the weights of temporal features to predict fluency scores
  • The interpretability issue (eXplainable Artificial Intelligence; XAI)
  • Latest model: 60% accuracy of CEFR level classification ($N = 84$)
Summary and future directions

The gap between SLA research and Language Testing in how fluency is defined and operationalized:

• SLA research has adopted a very narrow perspective to fluency, language tests may have a broader perspective, often combined with some other constructs (e.g., pronunciation, coherence).

• Only a few studies have investigated the construct of dialogic fluency.

A better understanding of fluency as a construct may offer insights into how to assess it automatically:

• Developmental patterns of fluency across proficiency levels
• Dynamics between fluency and conceptual and linguistic demands
• Speech elicitation techniques (esp. dialogue) for high ratability
Acknowledgements

Japan Second Language Acquisition Research Forum

New Energy and Industrial Technology Development Organization