Measuring functional adequacy in L2 and L1 writing across languages. Proposal of a rating scale

Ineke Vedder
ACLC (Amsterdam Center Language and Communication)
University of Amsterdam, The Netherlands

Standard and variation in second language education:
A cross-linguistic perspective
Roma Tre, 13-11-2015
Outline

1. Introduction: Standard, variation, functional adequacy (FA)
2. FA: definitions and assessment
3. FA in L2 writing: Study 1
4. Towards a new rating scale of FA
5. Testing out the new rating scale: Study 2
6. Pedagogical implications and conclusions
7. Next steps
1. Introduction: Standard, variation, functional adequacy (FA)
standard variations
Variation: Context

- Geographic, historical
  - macro (e.g. Amsterdam, Rome, 2015),
  - meso (e.g. speech community),
  - micro (e.g. classroom)

- Socio-pragmatic and linguistic:
  - Diglossia, dialects
  - Different languages
  - Source and target language (L1-L2)
Variation: Task

Task

- Type (e.g. narration)
- Task requirements, speech act (e.g. convincing)
- Modality (e.g. oral, written)
- Genre (e.g. recipee)
- Register (e.g. informal)
- Interlocutor (e.g. school children)
Variation: interindividual

- Individual learner characteristics
  - e.g. motivation, aptitude, extraversion

- Language proficiency
  - L1 and/or L2
  - basic, intermediate, advanced

- Instructional background

- …..
Functional Adequacy (FA)

- Pragmatically and linguistically *acceptable*

- Context-dependent: in given pragmatic and linguistic context

- Task-related: with respect to speech-act and language task

- Interactive: to be established by A (‘speaker’) and B (‘listener’)

---

11
Aims of presentation

- Investigation of the role of Functional Adequacy (FA) in assessing L2 proficiency within task-related framework (TBLT)

- Development and testing out of a rating scale for FA in L2 (and L1) writing
  - Corpus: written L2 and L1 data of Dutch, Italian (and Spanish)

Kuiken & Vedder 2014, forthcoming; Kuiken, Vedder & Gilabert 2010
Questions

- Componential nature of FA: underlying constituents of construct?

- How do we have to assess a multifaceted construct as FA?

→ Proposal of rating scale FA
Functional \((\text{in})\)adequacy: 2 examples
Functional (in)adequacy: two examples

- Related to socio-pragmatic context: Leaving a phone message to the dentist (L1)

- Related to language task: Giving advice to a friend (L2)
Phone message to the dentist (Italian L1; Dutch L1)
Ehm, mi dispiace, sono mortificata, ma purtroppo devo annullare l’appuntamento di oggi, perché non ho modo di arrivare da voi, non ho la macchina, e purtroppo, scusatemi, dovrei fissare un altro appuntamento. Se gentilmente mi può richiamare sul cellulare, perché sono al lavoro, grazie.
Goedemiddag, u spreekt met [+ name], ik wil een andere afspraak maken, ik had een afspraak op 4 december, telefoonnummer is nul zes vier drie nul één nul nul één twee en anders nul drieëntwintig vijf vijf zes acht twaalf drie, bedankt.
Caro Orlando,
Secondo gli tuoi condizione creo che il numéro 3, B&B Hotel Migani Spiaggia è il più appropriata. Non è vicino al centro, ma è vicino al boulevard, che anche è interessante. E di notte ha abbastanza movimento e locomozioni fino a tardi. Va per te, lo so sicuro!
Cara Valentina

Sono molto felice che hai scelto cinque diverse opzioni per andare in vacanza. Non te ho visto lungo. C’è una scelta difficile. Però tu ci sei per me, non c’è importante per me dove siamo. Ti amo. Ti vorrei abbracciare e non lasciare mai. Quindi penso che la prima scelta è ottima, perché il B&B assomiglia molto tranquillo. Inoltre c’è un giardino e fanno una prima colazione fantastica. Per di più abbiamo la tutto di tempo per goderci La vita e l’un l’altro. Ti desidero.

Un baciozzo, X.
2. Functional Adequacy: Definitions and assessment
FA and CAF: two dimensions of L2 proficiency

- Complexity, Accuracy and Fluency (CAF) usually considered as fundamental aspects of L2 proficiency

- Few studies report on the functional dimension of L2 output: Functional Adequacy (FA)

- Assessment of L2 proficiency is not possible without taking into account the FA of L2 production

Kuiken et al. (2010), Kuiken & Vedder (forthcoming), Pallotti (2009)
Definitions of FA in the literature

- Success of information transfer Upshur & Turner (1995)

- Socio-pragmatic appropriateness McNamara & Roever (2007)

- Intercultural communicative competence Hismanoglu (2011)

- Coherence and cohesion of text (Knoch 2009)

- Successful task fulfillment Alanen et al. (2010), De Jong et al. (2012), Kuiken, Vedder & Gilabert (2010), Kuiken & Vedder (2014, submitted)
Assessing FA

- No unanimity as to (i) how FA should be assessed; (ii) by means of which features it is determined

- Lack of general and quantitative measures to assess FA, comparable to CAF measures

- Assessment of FA by means of overall judgments on Likert scales Kuiken et al. (2010), Bridgeman et al. (2011), De Jong et al. (2012)

- Think-aloud protocols, panel discussions and interviews to investigate rater behaviour Brown (2007), Kuiken & Vedder (2014, forthcoming)
FA: Successful task fulfillment

- Interpersonal, task-related construct, involving two participants

- Focus on a specific task, to be carried out by A (and perceived by B)

- Quantity, relevance, manner and quality of the message transmitted by A (Grice 1975)

- Content, task requirements, comprehensibility, coherence/cohesion as important dimensions of FA
Rationale

- Need to assess L2 proficiency (also) in terms of functional adequacy
- FA as a multi-faceted construct
- Assessment of FA as a separate dimension from CAF
- Development of a (new) rating scale of FA
3. FA in L2 writing: Study 1
Objectives

- Assessment of writing proficiency in L2 (and L1) Dutch, Italian, Spanish: FA - Linguistic Complexity

- Linguistic features distinguishing levels of (written) performance in L2 (and L1)

- Investigation of rating behaviour and rater perceptions while assessing writing proficiency in L2 (and L1)

Participants, tasks and raters

- Participants
  - L2: 32 Dutch, 39 Italian, 23 Spanish, CEFR level: A2-B1
  - L1: 17 Dutch, 18 Italian, 10 Spanish

- Tasks
  - Two writing tasks (short argumentative essay)
  - C-test

- Raters
  - Expert raters (native speakers; L2 teachers)
  - Dutch: 4; Italian 3; Spanish: 3
  - Global Likert scales (1-6)
Research questions

- What are the linguistic and/or functional features which raters consider to be crucial?

- Which features do raters associate with a particular rating level?

- Do raters use different strategies when rating low versus high proficient learners?

- What do raters do when judging texts of L2 learners versus those written by L1 speakers?
Back to nature

*Back to nature* is an organization which tries to save the few authentic places that are still left in our country.

SPMC (Sports for the Physically and Mentally Challenged)

SPMC supports sport clubs which are willing to create facilities for handicapped children.

Animals first

*Animals first* takes care of animals which are in danger.
Task 1
Every year your university supports a charitable organization with an amount of € 50,000. The University Board has nominated three organizations:

1) Back to Nature,
2) SPMC (Sports for the Physically and Mentally Challenged),
3) Animals First.

Out of these three organizations one has to be selected. The selection is made by a Committee consisting of students. Every member of the committee has to write a report in which s/he states who the winner should be and why. On the basis of the arguments given in the reports the University Board will then decide who the winner will be.

This year you are on the Selection Committee. Read the brief descriptions of the organizations below. Determine which organization should receive the money this year. Write a text in which you give at least three arguments for your choice. Try to be as clear as possible and include the following points in your report:

to which organization the money should go;
what the aim of the organization is;
who will be helped by the organization;
why the university needs to support this organization (give three arguments)

You have 35 minutes to write your report and you need to write at least 150 words (about 15 lines). The use of a dictionary is not allowed.
General proficiency: C-test

Di queste parole manca la metà o la metà + 1. Completa le parole. Esempio: Domani pe- - - di and- - - al ci - - -

Cupido sì, amore eterno no

Terra di poeti, viaggiatori e innamorati. Il B - - Paese ris- - - il cu- - - dei gio- - - -? Pare pro- - - di s-. I gio- - - italiani so - - i p- - innamorati d- tutti. N- - credono a- - amore ete- - - , ma a- colpo d- fulmine s-. Nove ital- - - su di- - - hanno de- - - di ess- - - già st- - - innamorati. Ma solo un giovane su due crede all’amore eterno
### Scales and descriptors for the assessment of Functional Adequacy

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The information which the participant communicates</strong></td>
<td><strong>The information which the participant communicates</strong></td>
<td><strong>The information which the participant communicates</strong></td>
<td><strong>The information which the participant communicates</strong></td>
<td><strong>The participant provides complete information about</strong></td>
<td><strong>The participant provides very complete and precise information about</strong></td>
</tr>
<tr>
<td><em>sarcely describes</em> the organization, but does not describe its objectives or the people who will benefit from it.</td>
<td><em>briefly describes</em> the organization, and scarcely describes some of its objectives, or the people who will benefit from it.</td>
<td><em>briefly describes</em> the organization, its objectives, or the people who will benefit from it.</td>
<td><em>clearly describes</em> the chosen organization, its objectives, or the people who will benefit from it.</td>
<td>provides complete information about the organization, its objectives, or the people who will benefit from it.</td>
<td>provides very complete and precise information about the organization, its objectives, or the people who will benefit from it.</td>
</tr>
<tr>
<td>The arguments are not connected to the organization’s goals and beneficiaries.</td>
<td>The arguments are simple (i.e. single statements not supported by other arguments) and poorly connected to the organization’s goals and beneficiaries.</td>
<td>The arguments are simple (i.e. single statements not supported by other arguments) but connected to the organization’s goals and beneficiaries.</td>
<td>The arguments are complex (e.g. single or multiple statements supported by other arguments) and clearly connected to the organization’s goals and beneficiaries.</td>
<td>The arguments are complex and elaborate (e.g. single or multiple statements supported by other arguments) and clearly connected to the organization’s goals and beneficiaries.</td>
<td>The arguments are complex and elaborate, and also highly convincing.</td>
</tr>
<tr>
<td>And/or:</td>
<td>And/or:</td>
<td>And/or:</td>
<td>And:</td>
<td>And:</td>
<td>And:</td>
</tr>
<tr>
<td>The text lacks coherence, and it is very confusing.</td>
<td>The text lacks coherence, and it is confusing.</td>
<td>The text meets the minimum requirements.</td>
<td>The text is coherent.</td>
<td>The text flows smoothly, it is coherent, and is convincing.</td>
<td>The text flows smoothly. It is well constructed, coherent and convincing.</td>
</tr>
<tr>
<td>→ an unsuccessful contribution</td>
<td>→ a weak contribution</td>
<td>→ a moderately successful contribution</td>
<td>→ a successful contribution</td>
<td>→ a very successful contribution</td>
<td>→ a highly successful contribution</td>
</tr>
</tbody>
</table>
Results

- Interrater reliability (Cronbach’s α): L2: .72-.79; L1: .70-.90)

- Difficulty of raters to fit judgments of FA into necessary simple categories of the overall Likert scale

- Raters are partly led by linguistic features of the L2 text when assessing FA

4. Towards a new rating scale of FA
Requirements

- ‘Deconstruction’ of relevant components of FA
- FA descriptors independent from linguistic descriptors in terms of CAF
- ‘Objective’ and ‘countable’ scale descriptors
- Applicability for non-expert raters
- Possibility of use in both L2 and L1
Deconstructing FA
Grice (1975), Knoch (2007), Bridgeman et al. (2012), De Jong et al. (2012)

- Content
  - Is the number of information units (i.e. ‘ideas’ or ‘concepts’) provided in the text adequate and relevant?

- Task requirements
  - Have the requirements of the task been fulfilled successfully, with respect to text genre, register and speech?

- Comprehensibility
  - How much effort is required of the reader in order to understand text purpose and ideas expressed by the writer?

- Coherence and cohesion
  - Is the text coherent and cohesive (e.g. use of cohesive and anaphoric devices and strategies)?
Content: Is the number of information units provided in the text adequate and relevant?

1. The number of ideas is **not at all adequate** and insufficient and the ideas are unrelated to each other.
2. The number of ideas is **scarcely adequate**, the ideas lack consistency.
3. The number of ideas is **somewhat adequate**, even though they are not very consistent.
4. The number of ideas is **adequate** and they are sufficiently consistent.
5. The number of ideas is **very adequate**, they are very consistent to each other.
6. The number of ideas is **extremely adequate** and they appear very consistent to each other.
Have the Task Requirements been fulfilled successfully (e.g. genre, speech acts, register?)

1. **None** of the questions and the requirements of the task have been answered.
2. **Some (less than half)** of the questions and the requirements of the task have been answered.
3. **Approximately half** of the questions and requirements of the task have been answered.
4. **Most (more than half)** of the questions and the requirements of the task have been answered.
5. **Almost all** the questions and the requirements of the task have been answered.
6. **All** the questions and the requirements of the task have been answered.
Comprehensibility: How much effort is required to understand text purpose and ideas?

1. The text is not at all comprehensible. Ideas and purposes are unclearly stated and the efforts of the reader to understand it are ineffective.

2. The text is scarcely comprehensible, its purposes are not clearly stated and the reader struggles to understand the ideas of the writer. The reader has to guess most of the ideas and purposes.

3. The text is somewhat comprehensible, some sentences are hard to understand at a first reading. A second reading helps clarifying text purposes and the ideas conveyed, but some doubts persist.

4. The text is comprehensible, only sentences are unclear but are understood without too much effort after a second reading.

5. The text is easily comprehensible and reads smoothly, understandability is not an issue.

6. The text is very easily comprehensible and highly readable, the ideas and the purpose are clearly stated.
Coherence and cohesion: Is the text coherent and cohesive (e.g. cohesive devices, strategies?)

1. The text is **not at all coherent**. Unrelated progressions and coherence breaks are very common. The writer does not use any anaphorical device. The text is **not at all cohesive**. Connectives are hardly ever used and ideas are unrelated.

2. The text is **scarcely coherent**. The writer often uses unrelated progressions; when coherence is achieved, it is often done through repetitions. Few anaphorical devices are present. There are some coherence breaks. The text is **not very cohesive**. Ideas are not well linked by connectives, that are rarely used.

3. The text is **somewhat coherent**. Unrelated progressions and/or repetitions are frequent. More than two sentence in a row can have the same subject (even when the subject is understood). Some anaphorical devices are used. There can be few coherence breaks. The text is **somewhat cohesive**. Some connective are used, but they are mostly conjunctions.

4. The text is **coherent**. Unrelated progressions are somewhat rare, but the writer sometimes relies on repetitions to achieve coherence. Enough anaphorical devices are used. There might be some coherence breaks. The text is **cohesive**. The writer makes good use of connectives, sometimes not limiting it to conjunctions.

5. The text is **very coherent**: when the writer introduces a new topic, it is usually done by using connectives or connective phrases. Repetitions are very infrequent. Anaphorical devices are numerous. No coherence breaks. The text is **very cohesive** and ideas are well linked by adverbial and/or verbal connectives.

6. Writer ensures **extreme coherence** by integrating new ideas in the text with connectives or connective phrases. Anaphorical devices are used regularly. Few incidences of unrelated progressions and no coherence breaks at all. The structure of the text is **extremely cohesive**, thanks to a skillful use of connectives (especially linking chunks, verbals and adverbials) that are often used to describe relationships between ideas.
5. Testing out the new rating scale: Study 2
Research questions
Kuiken & Vedder (forthcoming)

- How do non-expert raters judge the functional adequacy of argumentative texts of L2 (and L1) writers of Dutch and Italian?

1. What are the *interrater reliability scores* on the 4 dimensions of FA?
2. How are raters’ judgments on the *4 dimensions of FA correlated*?
3. How are raters’ judgments in *L2 and L1 correlated*?
4. How are raters’ judgments in the *two tasks* correlated?
Raters

- Non-expert raters
  - Italian: 4; Dutch: 4
  - Native speakers, university students

- Use of texts of Study 1

- 2 training sessions for each language
1. Interrater reliability for Dutch and Italian
(Intraclass Correlation Coefficient)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Dutch</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>.841**</td>
<td>.838**</td>
</tr>
<tr>
<td>Task requirements</td>
<td>.824**</td>
<td>.725**</td>
</tr>
<tr>
<td>Comprehensibility</td>
<td>.940**</td>
<td>.901**</td>
</tr>
<tr>
<td>Coherence/Cohesion</td>
<td>.860**</td>
<td>.867**</td>
</tr>
</tbody>
</table>
2. Correlations between the four dimensions

<table>
<thead>
<tr>
<th>Dim.</th>
<th>Dutch</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cont.</td>
<td>.848**</td>
<td>.814**</td>
</tr>
<tr>
<td>T.R.</td>
<td>.694*</td>
<td>.851*</td>
</tr>
<tr>
<td>Compr.</td>
<td></td>
<td>.873**</td>
</tr>
</tbody>
</table>
3. L2 vs L1: Descriptives for L2 and L1

<table>
<thead>
<tr>
<th>Dim.</th>
<th>L2-L1</th>
<th>Dutch</th>
<th></th>
<th>Italian</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Cont.</td>
<td>L2</td>
<td>32</td>
<td>2.92</td>
<td>.46</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>17</td>
<td>3.96</td>
<td>.55</td>
<td>18</td>
</tr>
<tr>
<td>T.R.</td>
<td>L2</td>
<td>32</td>
<td>3.04</td>
<td>.59</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>17</td>
<td>4.04</td>
<td>.60</td>
<td>18</td>
</tr>
<tr>
<td>Com.</td>
<td>L2</td>
<td>32</td>
<td>3.09</td>
<td>.47</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>17</td>
<td>4.79</td>
<td>.18</td>
<td>18</td>
</tr>
<tr>
<td>C&amp;C</td>
<td>L2</td>
<td>32</td>
<td>2.84</td>
<td>.45</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>L1</td>
<td>17</td>
<td>4.16</td>
<td>.47</td>
<td>18</td>
</tr>
</tbody>
</table>
3. L2 vs L2: Differences between L2 and L1

<table>
<thead>
<tr>
<th>Dim.</th>
<th>Dutch</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>t</td>
</tr>
<tr>
<td>Cont.</td>
<td>47</td>
<td>-7.010</td>
</tr>
<tr>
<td>T.R.</td>
<td>47</td>
<td>-5.588</td>
</tr>
<tr>
<td>Com.</td>
<td>47</td>
<td>-14.370</td>
</tr>
<tr>
<td>C&amp;C</td>
<td>47</td>
<td>-9.584</td>
</tr>
</tbody>
</table>
4. Correlations between task 1 and 2

<table>
<thead>
<tr>
<th>Dim.</th>
<th>Dutch</th>
<th></th>
<th>Italian</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Correl.</td>
<td>Sig.</td>
<td>N</td>
</tr>
<tr>
<td>Cont.</td>
<td>49</td>
<td>.623**</td>
<td>&lt;.001</td>
<td>57</td>
</tr>
<tr>
<td>T.R.</td>
<td>49</td>
<td>.704**</td>
<td>&lt;.001</td>
<td>57</td>
</tr>
<tr>
<td>Com.</td>
<td>49</td>
<td>.877**</td>
<td>&lt;.001</td>
<td>57</td>
</tr>
<tr>
<td>C&amp;C</td>
<td>49</td>
<td>.719**</td>
<td>&lt;.001</td>
<td>57</td>
</tr>
</tbody>
</table>
6. Pedagogical implications and conclusions
Summary of results

- Assessment of FA by non-expert raters: ✓
- Interrater reliability: ✓
- Correlations between different scale dimensions: ✓
- Possibility of use in both L2 and L1: Higher scores in L1
- Comparison task 1 and task 2: ✓
Applicability of the dimensions

1. Content ✓
2. Task requirements ✓
3. Comprehensibility ✓
4. Coherence and cohesion ✓
Concluding remarks

- Construct and definition of FA
- Different subscales rather than one global scale
- Other dimensions of FA
Pedagogical implications

- Use by non-expert raters (and teachers) after few training sessions
- Possibility of use for judging L2 (and L1) writing in classroom setting
- More focused feedback to L2 learners
La pragmatica per l’italiano L2
Giornata di studio
5 marzo 2015
Aula Magna, via Filippo Re 8, Bologna
7. Next steps
Next steps

- Applicability of the scale for:
  - written and oral performance
  - different tasks
  - different types of learner
    - individual differences between learners
    - cfr. findings Pallotti (2015; Italian L1/L2; school children)
  - different languages
  - FA in relation to online pragmatic learning (LIRA)
Acknowledgements

- Sible Andringa, Raffaello Ibba, Loeka van der Kooij, Tinka Makaske, Irene Mistro, Flora Prota, Melle de Vries, Jonne Voshol (University of Amsterdam)

- Saskia Nijman (University of Groningen)

- Chiara Sale (University of Cagliari)

- Gabriele Pallotti (University of Modena and Reggio Emilia)
References


References (cont’d)


References (cont’d)


