Justice in practice
– what it means
and why it is our responsibility

Cecilie Hamnes Carlsen
ALTE 53rd meeting and conference, Gent, April 24-26, 2019
2015-16: Stricter immigration policy in Norway

Norwegian conservative coalition government 2015:

Measures to face the refugee crisis

Proposed language and knowledge of society requirements for permanent residency and citizenship as part of stricter immigration legislation

«purpose to make it less attractive to apply for asylum in Norway»
Language requirements, CoE-states, 2007
Language requirements, CoE-states, 2018
Lack of consistency > it is not really about language

Language requirements for citizenship (2018), CoE/ALTE-report

- B2: 4
- B1: 8
- A2 wrt./ B1 oral: 1
- A2 sp./ B1 list.: 1
- A2: 7
- A2 oral: 1
- A1: 1
- Unspecified: 10
- No requirement: 7

Language tests as gate-keepers (shibboleths)

(McNamara, 2005, Gysen, Kuijper & Van Avermaet, 2009; Shohamy & McNamara, 2009; Strik et al., 2011; Bruzos et al., 2018, Khan, 2018)
“As it is difficult to see why some countries should have higher requirements than others for the same need, these differences throw doubt on the argument that immigrants need the knowledge they are required to demonstrate in order to successfully integrate.”

(Böcker & Strik, 2011: 182)
“Any ethical question of the legitimacy of such tests is in fact a political question: does the politics of the setting justify the use of the test? Your answer will obviously depend on whose side you are on.”
McNamara, 2005:356
Justice questions

“We need to examine the ramifications of tests, their uses, misuses, ethicality, power, biases, and the discrimination and language realities they create for certain groups and for nations [...].”

(Shohamy, 2007:144)
Is justice language testers’ responsibility?

«Validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment […]» (Messick, 1989:13).
Is justice language testers’ responsibility?

<table>
<thead>
<tr>
<th>EVIDENTIAL BASIS</th>
<th>TEST INTERPRETATION</th>
<th>TEST USE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Construct validity</td>
<td>Construct validity + Relevance/utility</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CONSEQUENTIAL BASIS</th>
<th>Value implications</th>
<th>Social consequences</th>
</tr>
</thead>
</table>

«The consequential aspect of construct validity includes evidence and rationale for evaluating the intended and unintended consequences of score interpretation and use in both the short and long term […]» (Messick, 1996: 251)
Is justice language testers’ responsibility?

«The consequential aspect of construct validity includes evidence and rationale for evaluating the intended and unintended consequences of score interpretation and use in both the short and long term [...]» (Messick, 1996: 251)
Is justice language testers’ responsibility?

«We need to examine the ramifications of tests, their uses, misuses, ethicality, power, biases, and the discrimination and language realities they create for certain groups and for nations [...]. All these topics fall under the theoretical legitimacy of Messick’s (1994, 1996) work on the consequences and values of tests.» (Shohamy, 2007:144)
## Is justice language testers’ responsibility?

<table>
<thead>
<tr>
<th>Author</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spolsky (1980)</td>
<td>need to be concerned with <strong>test misuse</strong></td>
</tr>
<tr>
<td>Hamp Lyon (1989)</td>
<td>must be aware of the potential <strong>consequences</strong> of what they do</td>
</tr>
<tr>
<td>Bachman (1990, 2005)</td>
<td>«It is the <strong>ways in which we use tests</strong> that is at the heart of language assessment”.</td>
</tr>
<tr>
<td>Lynch (1997)</td>
<td>need to engage with the broader sense of validity, i.e. <strong>test ethics</strong></td>
</tr>
<tr>
<td>Shohamy (1997+)</td>
<td>need to be aware that tests are <strong>tools of power and control</strong></td>
</tr>
<tr>
<td>Wall (2000)</td>
<td>need to investigate test <strong>impact on teaching and learning</strong></td>
</tr>
<tr>
<td>McNamara (2005+)</td>
<td>need to be aware of the the <strong>social dimension</strong> of and <strong>values</strong> in language testing</td>
</tr>
<tr>
<td>Kunnan (2009)</td>
<td>need to investigate the <strong>wider context of testing</strong></td>
</tr>
<tr>
<td>Davidson &amp; Lynch (2012)</td>
<td>“Validity in educational assessment is <strong>advocacy</strong> on behalf of students».”</td>
</tr>
<tr>
<td>Kane (2013)</td>
<td>test users and test developers have a shared <strong>responsibility for test use</strong></td>
</tr>
<tr>
<td>Pulinx &amp; Van Avermaet (2015)</td>
<td>language tests used to decide who are «**true citizens»»</td>
</tr>
<tr>
<td>Chalhoub-Deville (2016)</td>
<td>a validity framework needs to encompass test <strong>consequences</strong></td>
</tr>
<tr>
<td>Bruzos, Erdicia &amp; Khan (2018)</td>
<td>language testers are responsible that their tests are <strong>adequate for their purpose</strong></td>
</tr>
</tbody>
</table>
“Those who make decisions about test use necessarily bear much of the responsibility […], but a test developer […] shares this responsibility. In particular, test developers who suggest that a test can be used in a particular way have an obligation to support the claims that they make” (Kane, 2013: 62).
Responsibility for justice – part of professional standards

ILTA Code of ethics
Principle 9
Language testers shall regularly consider the potential effects, both short and long term on all stakeholders of their projects, reserving the right to withhold their professional services on the grounds of conscience.

EALTA guidelines for good practice
3. CONSEQUENCES
What use is made of the results?
What are the consequences of the assessment procedures for classroom practices?
What are the consequences of the results of the assessment for learners?

ALTE code of practice
Define what each examination assesses and what it should be used for.
Describe the population(s) for which it is appropriate.
So…

If justice is language testers’ responsibility, what does it mean in practice?
Language activism

“Once awareness of this process is evident, there is a need to engage in linguistic activism.” (Shohamy, 2006:xv)
Language activism

“Language activism is energetic action focused on language use in order to create, influence and change existing language policies” (Combs & Penfield, 2018:462)

“[…] stakeholders are seen as activists if they choose to use their position to raise awareness and call for policy changes” (Combs & Penfield, 2018:471).
Messick’s definition of validity gives us a licence to take action against injustice

It gives us a responsibility to act

What we consider injust, however, depends on our values
Language testing activism – justice in practice

- examples from Norway

LT policy
LT public debate
LANGUAGE TESTING ACTIVISM
LT development
LT research

Cecilie Hamnes Carlsen, Western Norway University of Applied Science
1. Language testing activism > policy
1. Language testing activism > policy

› Regional policy proposal – language requirement for labour
  › Non-native speaking kindergarten assistants
  › B2 requirement in all four skills

› Policy claim
  › Necessary in order to secure minority children’s Norwegian skills

› Why is it hard to justify?
  › only 5% of 20 000 test-takers gets B2 in the 4 skills
  › potentially harmful consequences for migrant workers (women in particular)
  › ignores the value of plurilingualism and multilingualism
  › ignores minority children’s need to communicate in a language they know
1. Language testing activism > policy

Action

› Responds to public hearing
› Arranged meeting with policy makers to inform about:
  › the CEFR, proficiency levels and learner profiles
  › B2-level being an academic level
  › the number and profile of candidates who receive B2 in all four skills
  › the value of plurilingualism and multilingualism
  › the importance of L1 in language learning
  › the potentially negative consequences of a B2-requirements on individuals and society
1. Language testing activism > policy

Result

› Policy makers withdrew the B2-proposal
› Introduced B1-requirement
   › One could still question the justification of B1-level in writing for kindergarten assistants
     › How much do kindergarten assistants write as part of their job?
     › What kinds of texts do they need to write?
   › Needs-analysis is necessary
2. Language testing activism > public debate

LT policy

LT development

LANGUAGE TESTING ACTIVISM

LT public debate

LT research
2. Language testing activism > public debate

› Policy proposal – requirements for integration/immigration
  › Language and knowledge of society (KoS) requirements
    › Permanent residency: A1 oral + written KoS-test in minority language
    › Citizenship: A2 oral + KoS-test in Norwegian

› Policy claim
  › «to make it less attractive to apply for asylum in Norway» (www.regjeringen.no)

› Why is it hard to justify?
  › language and KoS-learning is not equally easy for all learners
  › such requirements are particularly harmful for vulnerable groups (low-literates, refugees)
  › a danger that requirements get stricter once the law is implemented
  › may lead to segregation rather than integration (Bruzos, Erdocia & Khan; Böcker & Strik, 2011)
  › a validity problem – it is not really about language
2. Language testing activism > public debate

Action

› Respons to public hearing
› Active participation in public debate meetings
› Newspaper articles informing the public opinion about:
   › the CEFR-levels
   › KoS-test as implicit language and literacy requirement
   › the potential dangers for individuals and society
   › the experience from other countries
   › research results on impact of such requirements
2. Language testing activism > public debate

Action

› Response to public hearing
› Part-taking in
› Newspaper and other media
› CEFR-levels
› Kos-test as implicit language and literacy requirement
› Potential dangers for individuals and society
› Experience from other countries
› Research results on impact of such requirements

Når tester blir politikk

Med skjulerpede krav reknar man innvandrere med itle skolebaserns permanent opphold og familiegrunnloven. 

Kjendisse: Cecilie Hamnes Carlsen, Western Norway University of Applied Science

16 February 2016 19:30 - 21:00 | Åpen mote

Lista til Listhaug

Lista til Listhaug - hva betyr den egentlig?

Som svar på oknurret i avsløringen representerte innvandringsminister Sylvi Listhaug (FrP) i riksdagen en liste med saker som bidra til å stramme inn innvandrarkriteriene i Norge. IMER Bergen og debattører om innstramningssaker legger praksismedisinsk perspektiv.

Hvis lister over forlegging blir gjennomført vil den gjøre Norge til et hardere land for innvandrere. Forslagene inkluderer innsetting av familiefenomena, at bruk av misfattet opphold, krav til høyere språknivåer og at høyere beklagelse prøver at omsett finansieringsfordelene for å få permanent innvandrere.

For mange kan forleggingslinjer til takt virke abstrakte. Hva betyr egentlig innstramningssaker i praksis?

IMER Bergen og Christian Michelsons Institutt iustere

arrangement der fire eksperter gir innblikk i innstramningssaker. Du kan melde din interesse eller spre ordet om event.

Result

› Majority of hearing responses were negative to the proposed requirement
› Yet – politician chose to introduce language and KoS-requirements

› 2019 – new proposal to increase the level from A2 to B1 for citizenship (regjeringsplattformen)
› New public hearing and new battles to fight
3. Language testing activism > research
3. Language testing activism > research

› Language requirement for university entrance
  › Foreign students need to document their Norwegian skills
  › Several different tests can be used to meet the requirement
  › Admission requirements for the different tests are decided by policy makers
  › Language test scores are recoded to the same scale (1-6) for calculation of credit points

› Claim
  › The different requirements are comparable
  › Hence the entrance requirements can be recoded into the same point on the 1-6 scale

› Why is it hard to justify?
  › The different requirements do not appear to be aligned
  › It is likely that candidates at the same level of proficiency do not get the same credit
  › Potential discrimination of some candidates
3. Language testing activism > research

Action

› Arranged meeting with policy makers
› Explained why the current situation was potentially unfair
   › Qualitative comparison of assessment criteria and level descriptions
   › Prior correlation studies

› Agreement to conduct new study
   › Correlation of university admission requirements (three tests of Norwegian)
   › Aim: to find out if the recoding to same scale was fair
   › Same students taking several exams > compare scores
   › External research company did analyses and wrote report
Norskprøven for voksne innvandrere  
(Standardized test developed by Skills Norway)  
Entrance requirement: highest level measured  
B2 (A1 to B2-scale)

«Can write a clear, coherent text which is easy to follow. Can structure a text well by using connectors and other text-binding elements to create textual cohesion and progression»

«Has a broad vocabulary regarding both words and expressions. Can express him/herself with variation and a high degree of lexical precision. Errors occur but do not lead to misunderstandings»

«Has a good command of basic grammatical structures – both word inflection and sentence patterns. Good variation in sentence types. For the most part successful use of complex sentences. Errors occur but do not hinder communication»

Trinn 3-eksamen  
(Locally developed university entrance test)  
Entrance requirement: lowest pass grade  
E (F to A scale)

«Uses few and simple devices to create text binding»

«Vocabulary is limited, but sufficient»

«Syntax, morphology and orthography deviates considerably from target language norm, but the text is generally comprehensible»
3. Language testing activism > research

<table>
<thead>
<tr>
<th>New scale</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trinn 3-eksamen</td>
<td></td>
<td>E and D</td>
<td>C</td>
<td>B</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Norskprøven for voksne innvandrere</td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test i norsk-høyere nivå</td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Samordna opptaks table for recoding grades on different tests of Norwegian into the same 1-6 grade scale (complete scale available in Norwegian at: https://www.samordnaopptak.no/info/om/lover-og-regler/rangering/index.html)
3. Language testing activism > research

Design of the study

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (n = 43)</td>
<td>Group 1 (n = 43)</td>
<td></td>
</tr>
<tr>
<td>Group 2 (n = 30)</td>
<td></td>
<td>Group 2 (n = 30)</td>
</tr>
<tr>
<td>Group 3 (n = 49)</td>
<td></td>
<td>Group 3 (n = 49)</td>
</tr>
</tbody>
</table>
### 3. Language testing activism > research

<table>
<thead>
<tr>
<th>Trinn 3-eksamen</th>
<th>Norsk-prøven</th>
<th>Pass the admission requirement</th>
<th>Fail the admission requirement</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>7 (16 %)</td>
<td>30 (72 %)</td>
<td>37 (88%)</td>
<td></td>
</tr>
<tr>
<td>Fail</td>
<td>0 (0 %)</td>
<td>5 (12%)</td>
<td>5 (12%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>7 (16%)</td>
<td>35 (84%)</td>
<td>43 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

3. Language testing activism > research

Result

Clear recommendations from researchers:

› B2 (Norskprøven) does not correlate with grade E (Trinn 3-eksamen)
› B2 (Norskprøven) should not be equated to E and recoded into the same grade on the 1-6-scale (grade 2)
› B2 (Norskproven) correlates with grade B (Trinn 3-eksamen) and should be recoded into grade 4

› **Change in policy!**
3. Language testing activism > research

<table>
<thead>
<tr>
<th>New scale</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trinn 3-eksamen</td>
<td></td>
<td>E and D</td>
<td>C</td>
<td>B</td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Norskprøven for voksne innvandrere</td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test i norsk-høyere nivå</td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Samordna opptaks table for recoding grades on different tests of Norwegian into the same 1-6 grade scale (complete scale available in Norwegian at: [https://www.samordnaopptak.no/info/om/lover-og-regler/rangering/index.html](https://www.samordnaopptak.no/info/om/lover-og-regler/rangering/index.html))
3. Language testing activism > research

<table>
<thead>
<tr>
<th>New scale</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trinn 3-eksamen</td>
<td></td>
<td></td>
<td>E and D</td>
<td>C</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Norskprøven for voksne innvandrere</td>
<td></td>
<td></td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test i norsk-høyere nivå</td>
<td></td>
<td></td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Samordna opptaks table for recoding grades on different tests of Norwegian into the same 1-6 grade scale (complete scale available in Norwegian at: [https://www.samordnaopptak.no/info/om/lover-og-regler/rangering/index.html](https://www.samordnaopptak.no/info/om/lover-og-regler/rangering/index.html))
Endelig uttelling for gode resultater på norskprøven


Kompetanse Norge la ned mye arbeid høsten 2016 for at norskprøvekandidater skal få uttelling for sine resultater.


Kompetanse Norge og Samordna oppfikk ble enige om å få gjennomført en korrelasjonsundersøkelse av opptakskravene i norsk, der Norskkvoten for voksne innvandrere, Test i norsk - høyere nivå og Trinn 3-eksamen inngikk.

Resultatene av undersøkelsen viste blant annet at:
4. Language testing activism > test development

- LT policy
- LT public debate
- LT development
- LT research

LANGUAGE TESTING ACTIVISM
4. Language testing activism > test development

› Minority childrens’ language rights in school
  › Special curriculum in Norwegian as a second language
  › Subject teaching in their L1
  › L1-tuition
    › Only until they have «sufficient proficiency» in Norwegian to follow ordinary classes
    › There is no standardized assessment tool available

› Claim
  › Teachers can assess what is «sufficient»

› Why is it hard to justify?
  › Teachers lack in SLA and LT-competence
    › childrens’ language proficiency is often overestimated because of good pronunciation
  › The assessment is practiced very differently from school to school
  › Many minority children do not get what they need to succeed
4. Language testing activism > test development

Action

› Develop a test to assess what is «sufficient»

› Collaboration between Hordaland county municipality & Western Norway University of Applied Science
  › CEFR-based test of speaking, writing and reading
  › Tasks and assessment criteria developed with teachers
  › Systematic training of teachers in the county municipality > assessment literacy
  › Test for both summative and formative purposes
Vurderingsverktøy for kartlegging av norskferdigheter - for minoritetsspråklige elever i vgs

(Assessment tool for minority pupils in upper secondary school)
Summing up the line of argument presented

- Validity is in the core of language testers’ responsibility
- Messick’s concept of validity encompasses justice
- Hence justice is language testers’ responsibility
- Taking the responsibility for justice seriously, implies taking action
- This action needs to take on different forms to be efficient

- Examples of language testing activism in four contexts in Norway
  - LT policy (> labour market)
  - LT public debate (> permanent residency and citizenship)
  - LT research (> university admission)
  - LT development (> minority children’s language rights in school)


