



Bias for best in test delivery - minimising construct irrelevance, enhancing test fairness

Hye-won Lee, Cambridge University Press &
Assessment
Emma Bruce, British Council

EALTA, Belfast June 8 2024



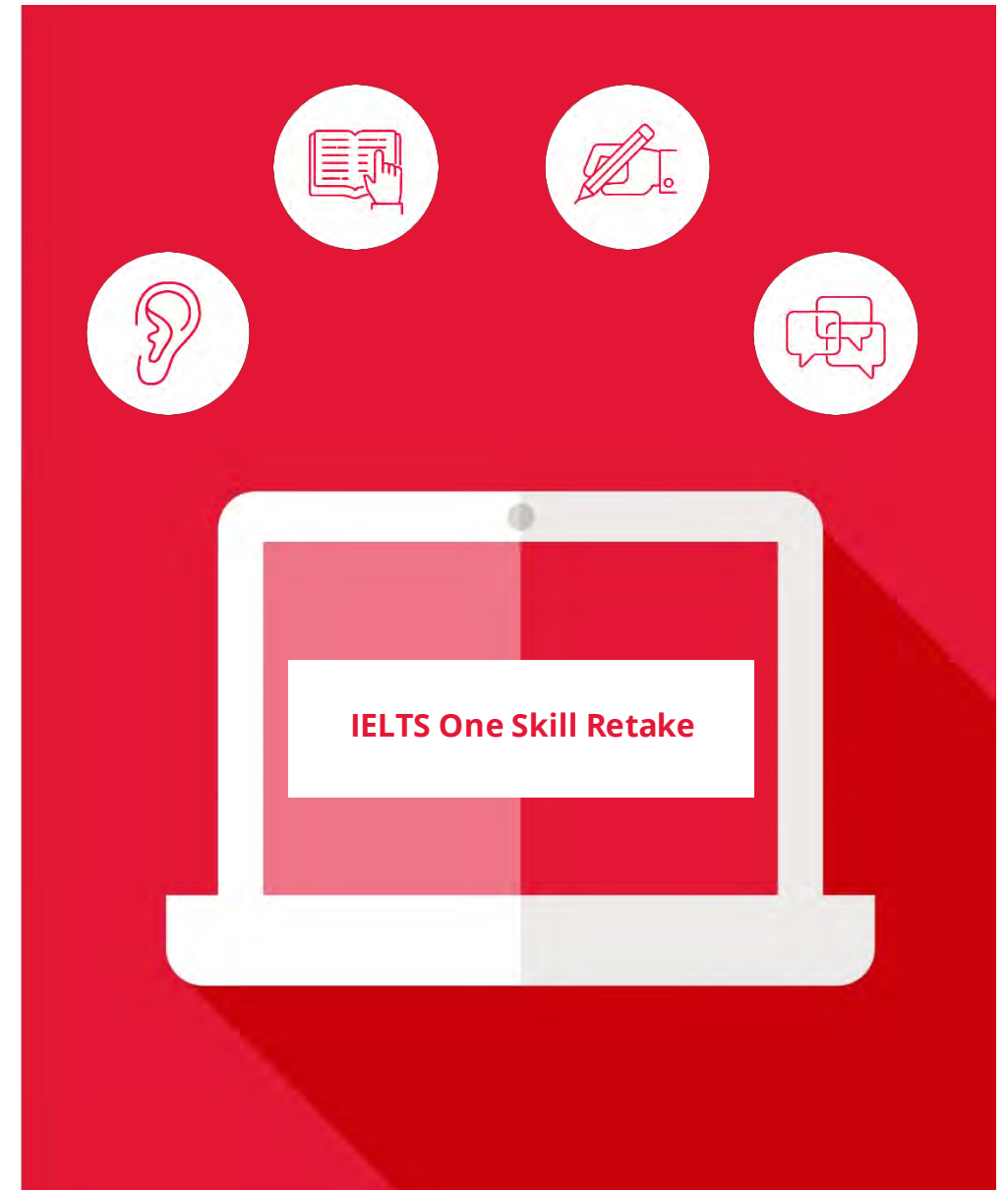
IELTS



IELTS One Skill Retake

Background

- Bias for best approach
- Measures in place to ensure standard of assessment is maintained:
 - One skill
 - 60-day window
 - Beneficial washback of focused study
 - Large item bank
 - Unique test taker identifier



An example test taker score profile

| | L | R | W | S | O |
|-----------|------------|------------|------------|------------|-----|
| 1st July | 7.0 | 7.0 | 7.0 | 6.5 | 7.0 |
| 8th July | 7.0 | 7.0 | 6.5 | 7.0 | 7.0 |
| 15th July | 7.0 | 6.5 | 7.0 | 7.0 | 7.0 |
| 22nd July | 6.5 | 7.0 | 7.0 | 7.0 | 7.0 |

- Some Test Score Users set skill-level requirements.
- This means a test taker can be defined by their lowest score in a single skill.
- See the example above which shows what a test taker's score could be if they were to take the test in quick succession.
- Variability between scores on individual sections is not unusual.

Possible scenarios for IELTS



Enquiry on Results (EOR)

- Test takers request a re-mark of one or more sections
- No later than six weeks after the test date
- Fee refunded if the score is increased

IELTS One Skill Retake

- Test takers retake one section (one skill once per original test)
- Within 60 days of the original test
- Cost is around two thirds of full test cost

Retake a Full Test

- Test takers retake all four sections
- No official restrictions on the number of retakes
- No mandatory waiting period before retake

Review of relevant literature

Retaking in educational assessment

- **Fairness:** Possibility of **error** in initial assessment due to such **extenuating circumstances** as illness or anxiety (Hausknecht et al., 2007; Lievens et al., 2005)
- **Typical outcomes: Improvement** on retests attributable to some combination of greater familiarity with the instrument and a self-selection effect (Hausknecht et al., 2007; Steiner, 2001)
- **Interpretation of retaking: *One* retest** to be justified, just in case the person was **unlucky** and received a score that reflected **a large negative error** (Roszkowski & Spreat, 2016)

Repeat test-taking / test retest in language assessment

- Repeated failure to meet specific sub-score requirements creates a **vicious cycle** and aggravates **frustration and anger** among some test takers (e.g., Hoang & Hamid, 2017)
- **Little research or data** available on numbers, attributes, and performance of repeaters of English language proficiency tests (Barkaoui, 2017; Green & Van Moere, 2020)
- Need to examine **what repeat test-taking presents** for validity arguments of L2 proficiency tests

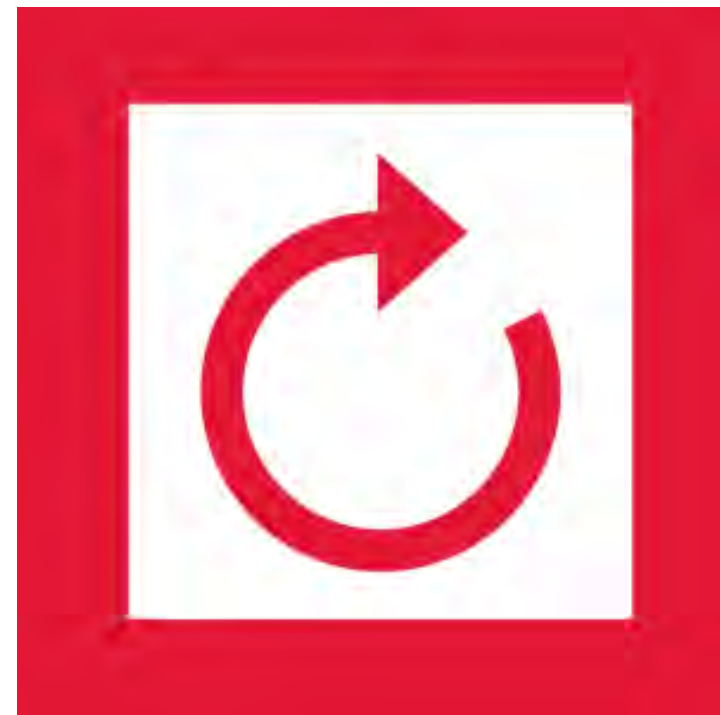
Test fairness

- Avoidance of the effects of any *construct-irrelevant factors* on the entire testing process (Walters, 2022)
 - E.g., Following the socio-cognitive framework (Weir 2005), O'Sullivan & Weir (2011): subdivide characteristics of the test taker into physical/physiological, psychological, and experiential
- IELTS One Skill Retake promotes positive effects of retesting by reducing construct-irrelevant variance
 - E.g., due to anxiety, unfamiliarity with the test and/or testing situation (Lievens et al., 2005)
- Kunnan's (2018) Fairness and Justice Argument
 - Fairness Sub-claim 4: The assessment provides appropriate **access**
 - Justice Sub-claim 1: The assessment has **beneficial consequences** to the test-taking community

Research into IELTS One Skill Retake

Research aims

- Identify **characteristics of test takers** who choose to take IELTS One Skill Retake in terms of demographics and patterns of full IELTS and IELTS One Skill Retake tests
- Examine **what contributes** (or does not contribute) **to variability** in IELTS One Skill Retake scores
- Understand the **perception** of test takers and recognising organisations to IELTS One Skill Retake



Methodology

| Data | Sample size | Analysis |
|--|-------------|---|
| Live database of all One Skill Retake test takers (e.g. scores, skills) | $n = 20K+$ | Descriptive statistics |
| Test taker questionnaire (Aug 2023 onwards) | $n = 507$ | Descriptive statistics Thematic analysis |
| Test taker interviews <ul style="list-style-type: none">• Pilot (Oct 2023)• Main study (Jan – May 2024) | $n = 27$ | Thematic analysis |
| Interviews with receiving organisations (ongoing) | $n = 4$ | Thematic analysis |

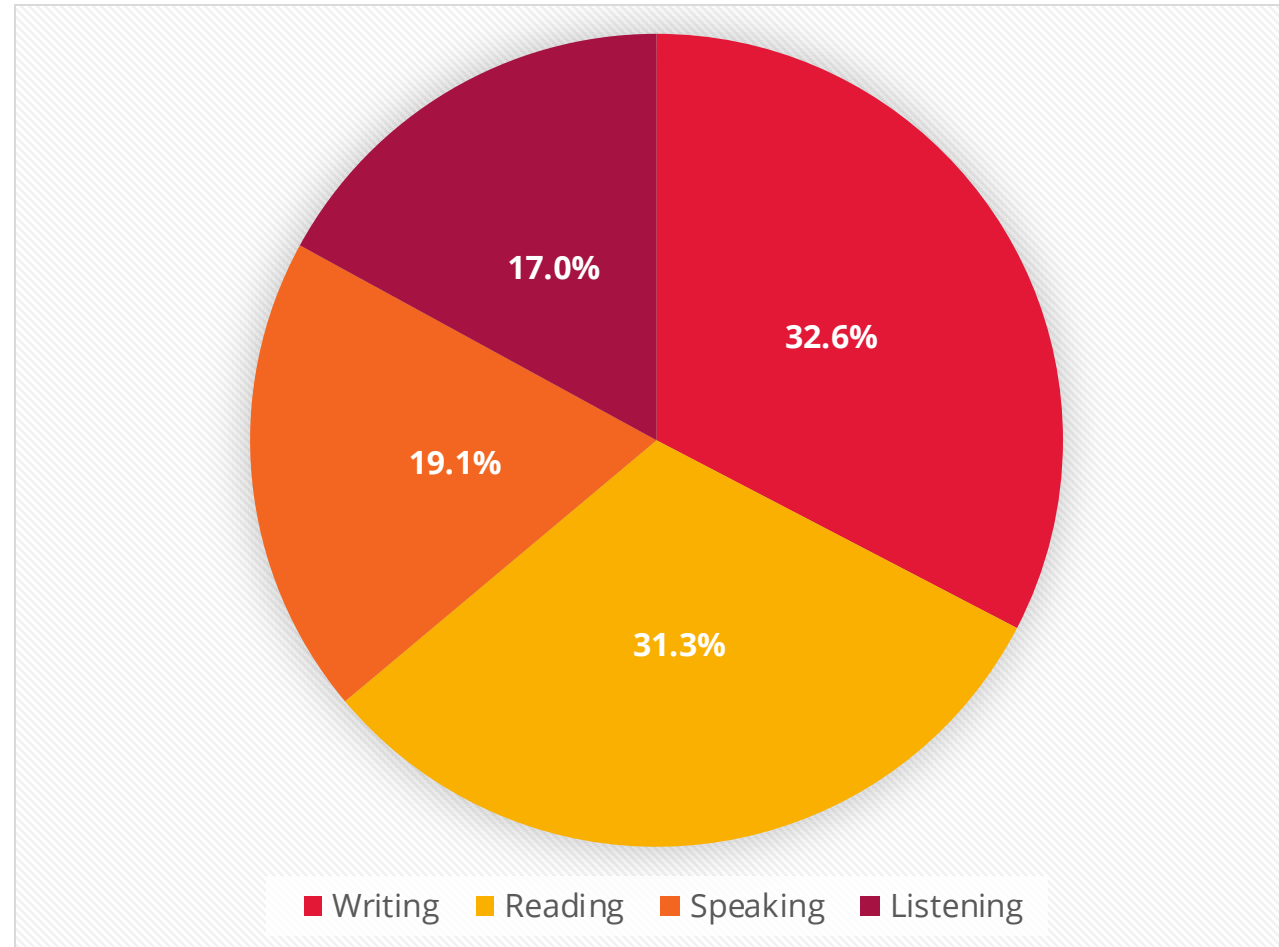
Test taker data

Demographics

- Tests sat in nearly 20 different countries
- 31.6% female, 37% male, 31.4% prefer not to say
- 60.4% were 26+ years old, no minors

Pattern of retaken components

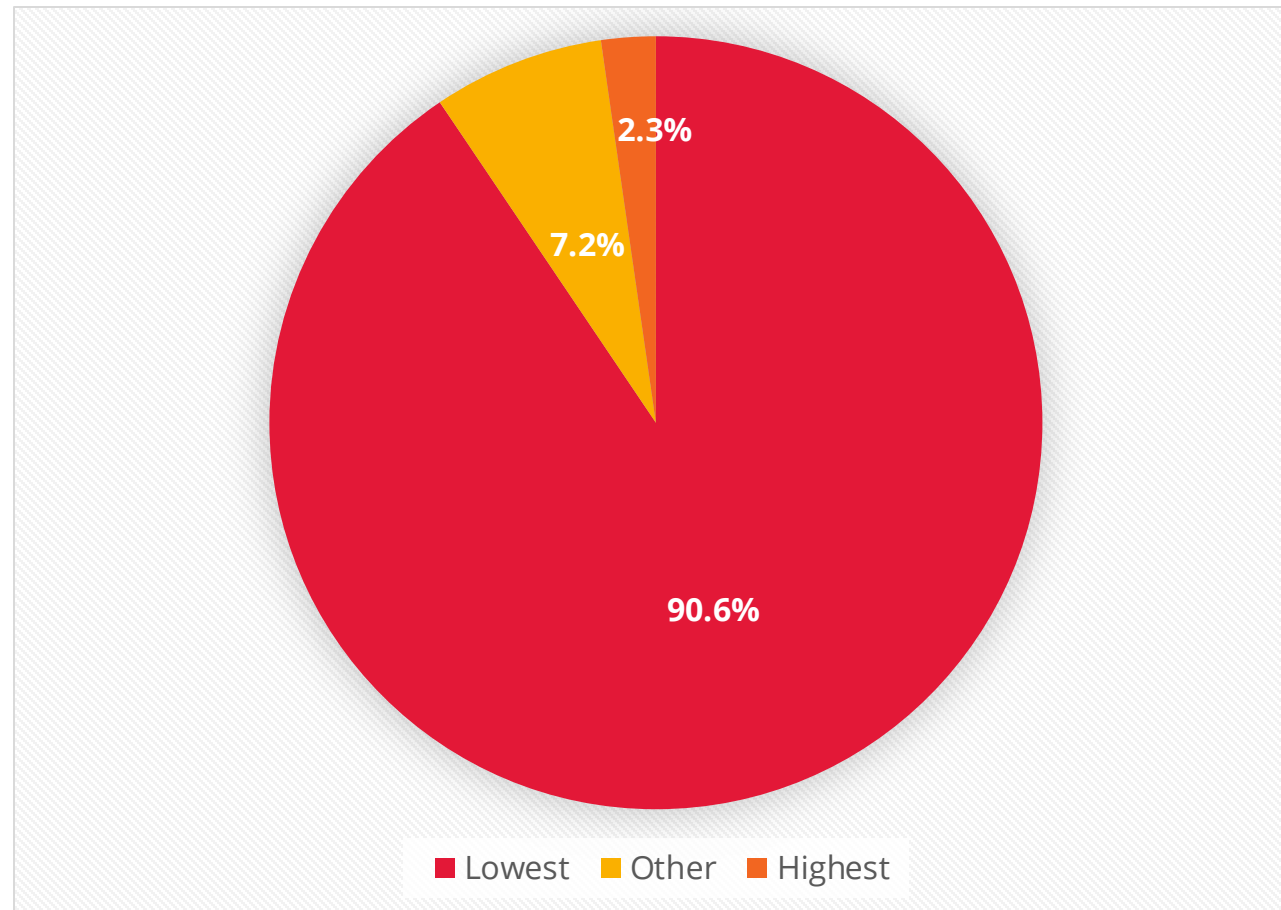
By Skill



n = approx. 20K

Pattern of retaken components

By Original Component Score



n = approx. 20K

Score change by skill

- 16% – 30% of OSR test takers see **no score change** in their chosen skill
- 52% – 56% of OSR test takers see **an increase of half to one band** in their chosen skill
- The **receptive skills see a slightly greater score gain**, especially Listening

n = approx. 20K

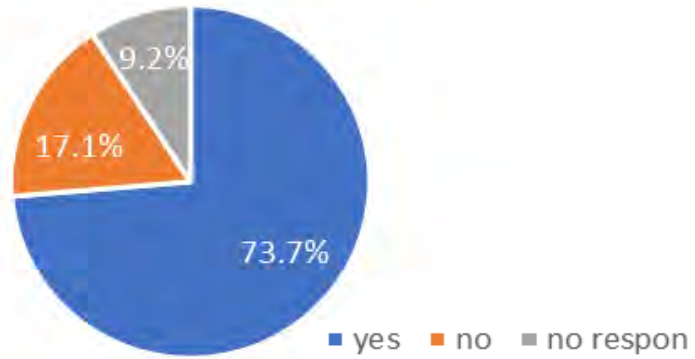
Overall score change

- 50% - 75% of OSR test takers see **no change in their overall band score** depending on which skill they retook
- 20% to 50% of OSR test takers see **an increase of half a band in their overall band score** depending on which skill they retook
- Average overall band score change for all OSR test takers is **not statistically significantly different** from that of full test repeaters

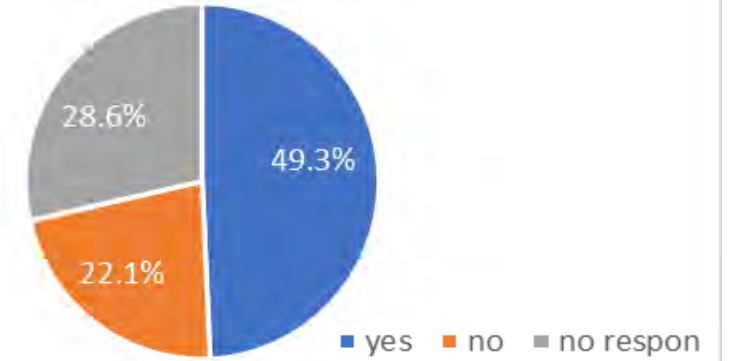
Test taker questionnaire and interview data

Did you achieve your desired score in your OSR?

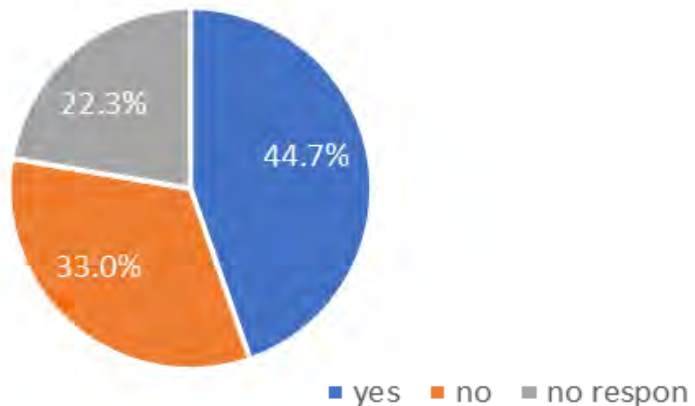
Listening



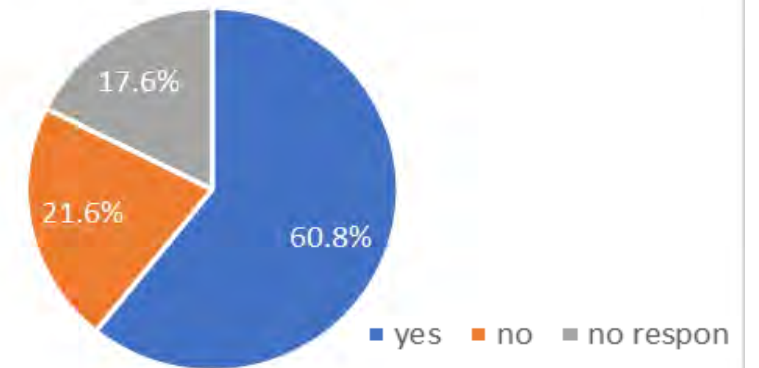
Reading



Writing



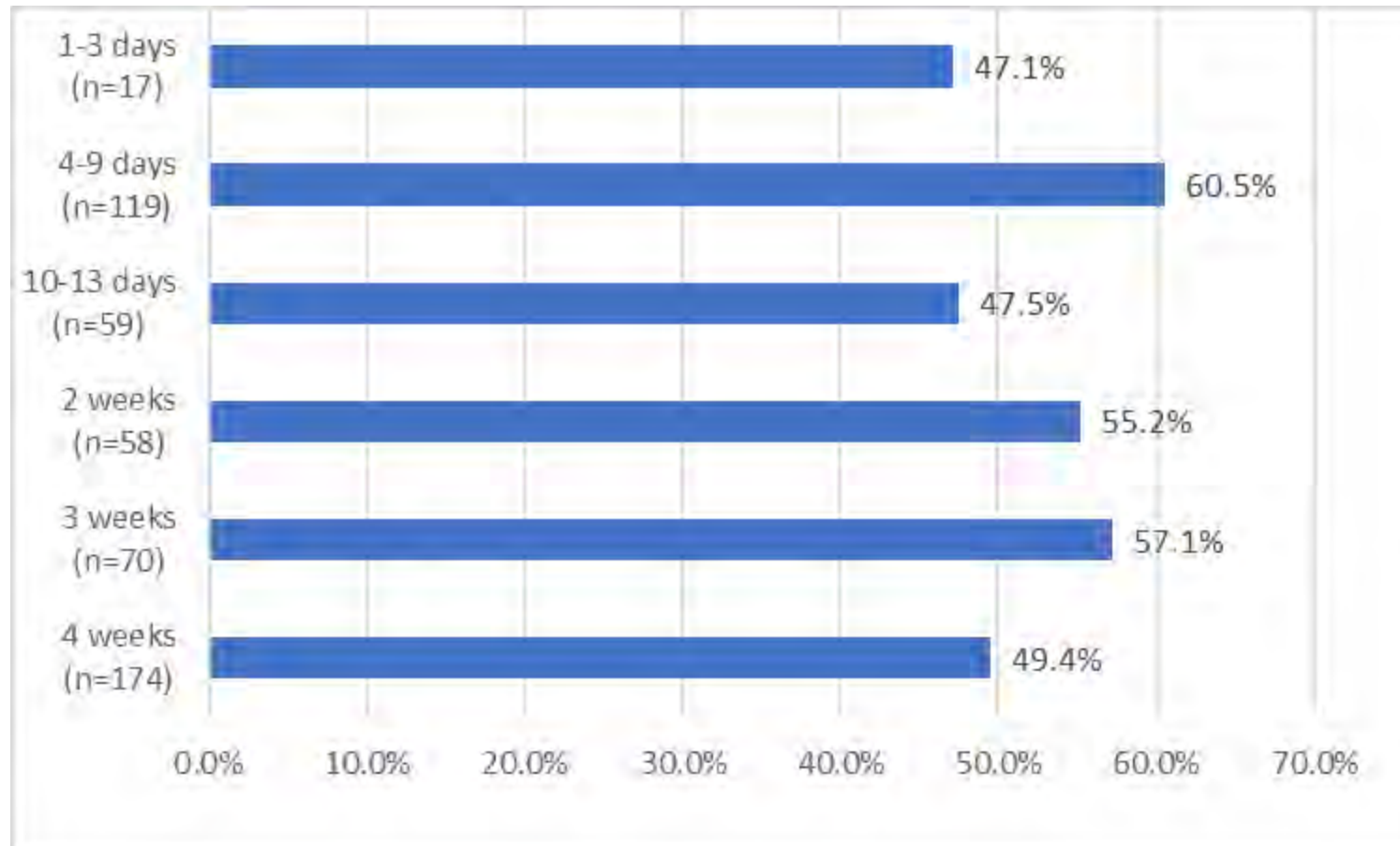
Speaking



n = 507

Did you receive your required score in your OSR?

Time period
b/w full and
OSR test



Percentage of Test Takers who received their required result

Time between the full IELTS and OSR test seemed to have little effect on test takers achieving their desired score.

n = 497

Why didn't you receive your desired score on the full test?

| Top responses | freq. |
|--|-------|
| Lack of practice / preparation | 73 |
| Time management | 44 |
| Lack of test familiarity | 32 |
| Stress / confidence | 26 |
| Unfair examiner (perception) | 21 |
| Fatigue / lack of focus | 19 |
| Familiarity w computer-based format | 13 |
| Distraction (e.g. by test environment) | 10 |

Optional open-ended question

Why didn't you receive your desired score initially?

I was **unable to practice it enough** that because there are four modules and I'm a working lady, I did not get enough time to work on each of the module ... a time constraint was one of the major thing, because **I could have achieved Band 8. But again, I needed to practice it more and I did not practice and I could not manage.**

Interviewee R - Writing

I got **really nervous** and it's, that's why I made so many mistakes in my previous exam.

Interviewee K - Writing

That is my mistake because when I get the test I was **sick at the time** and also I **don't show my performance.**

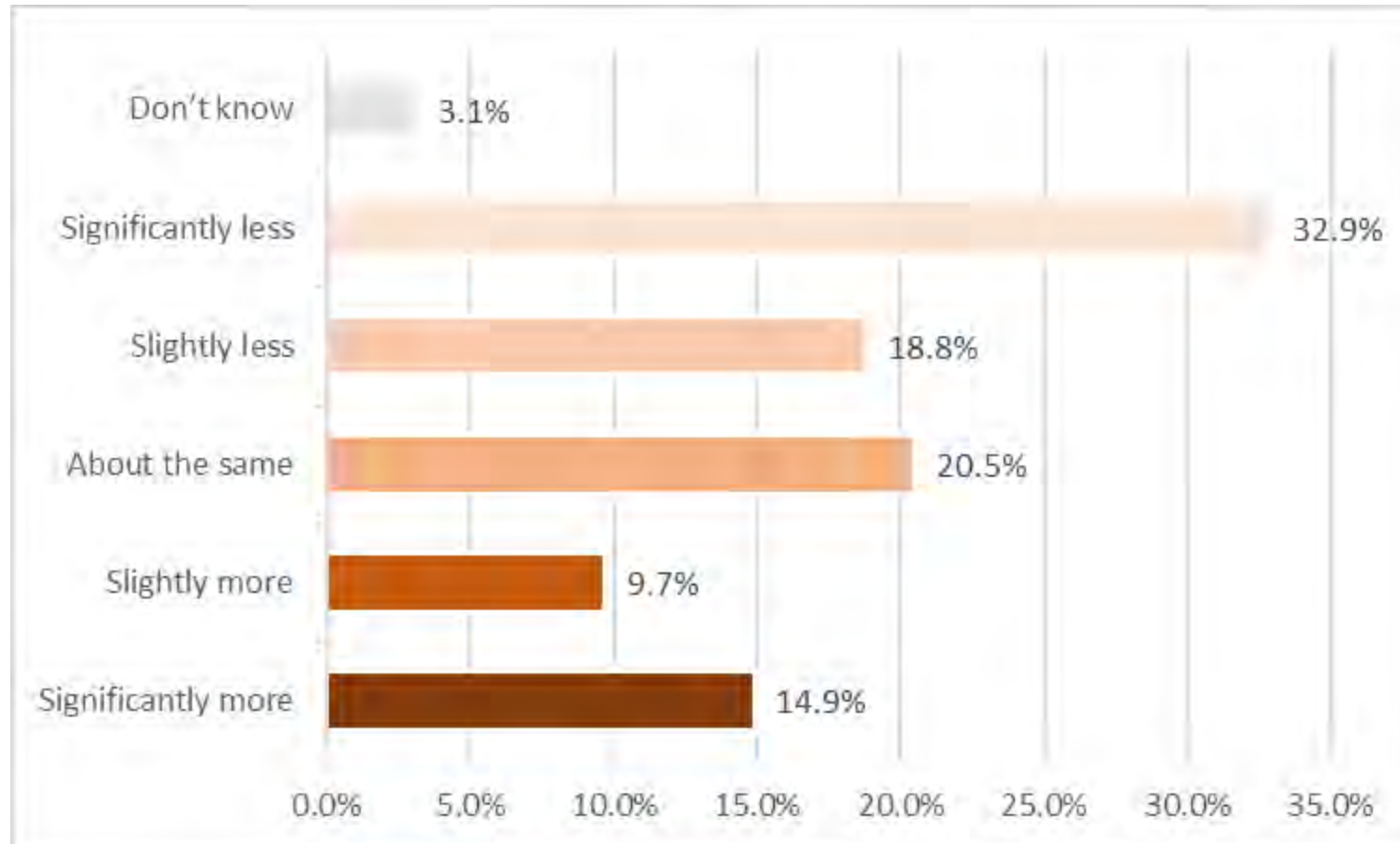
Interviewee O - Speaking

Before the 1st test, I was, I would say **50% aware of the whole test criteria.** A particularly maybe I was **taking things a little too lightly.** That's how a normal human being would react if he or she hasn't been experiencing the same scenario.

Interviewee J - Writing

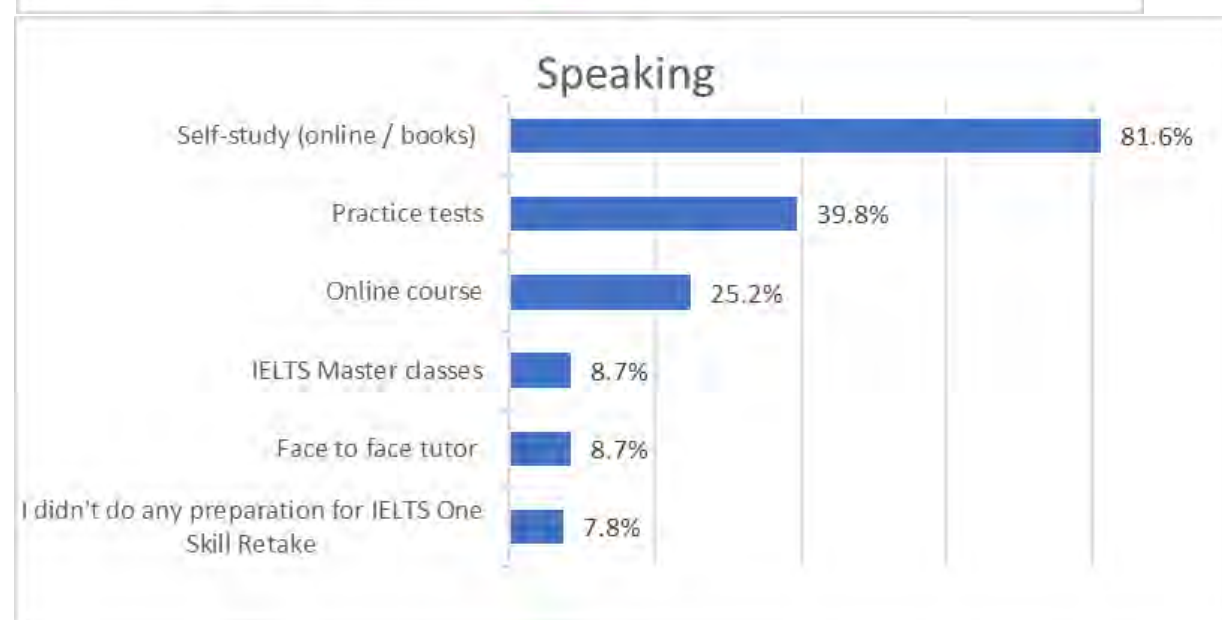
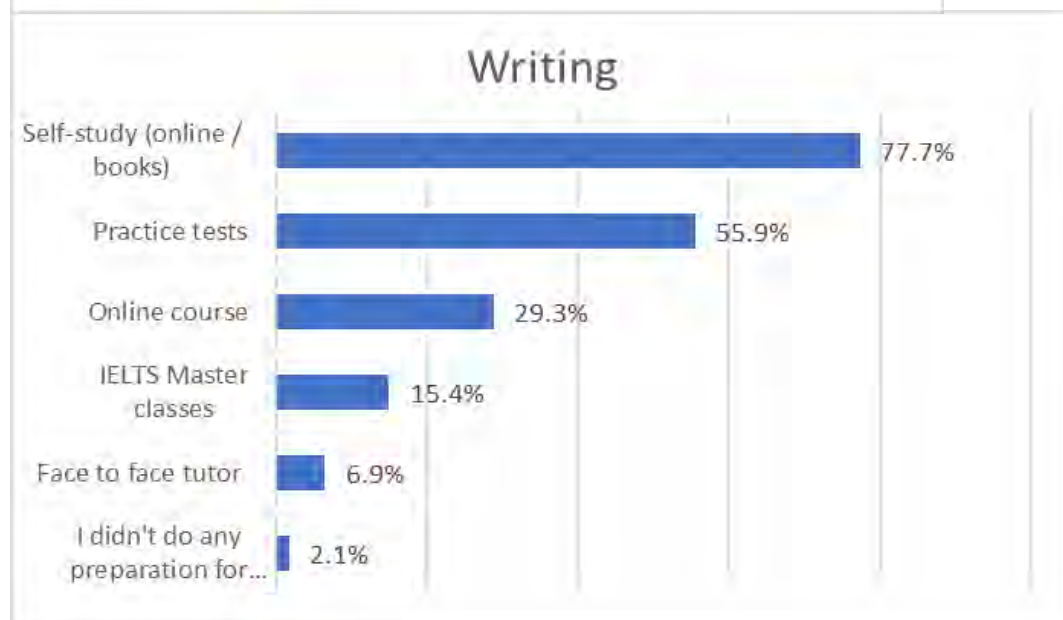
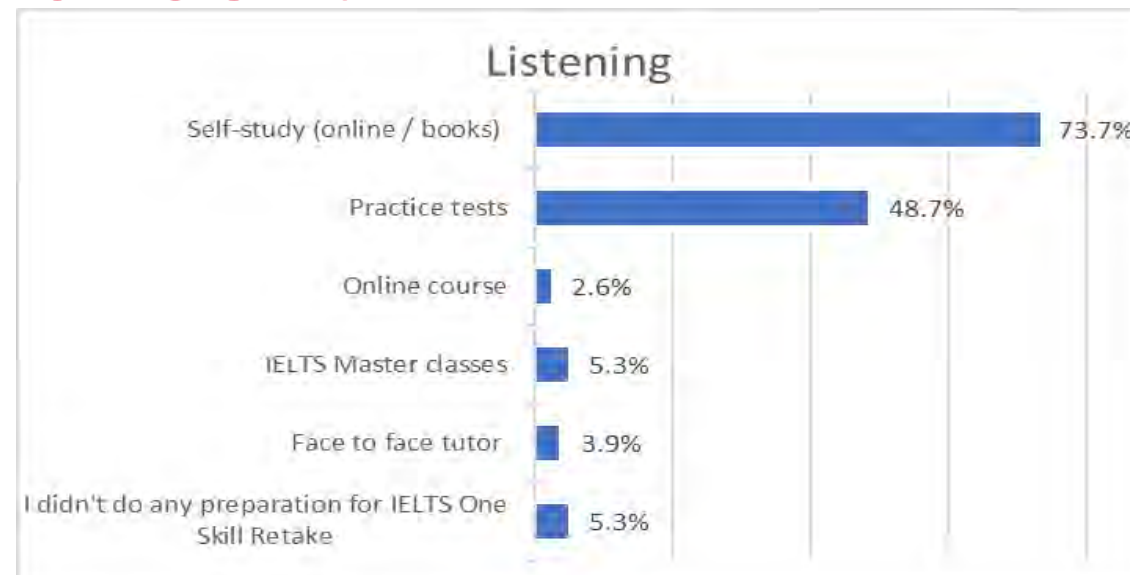
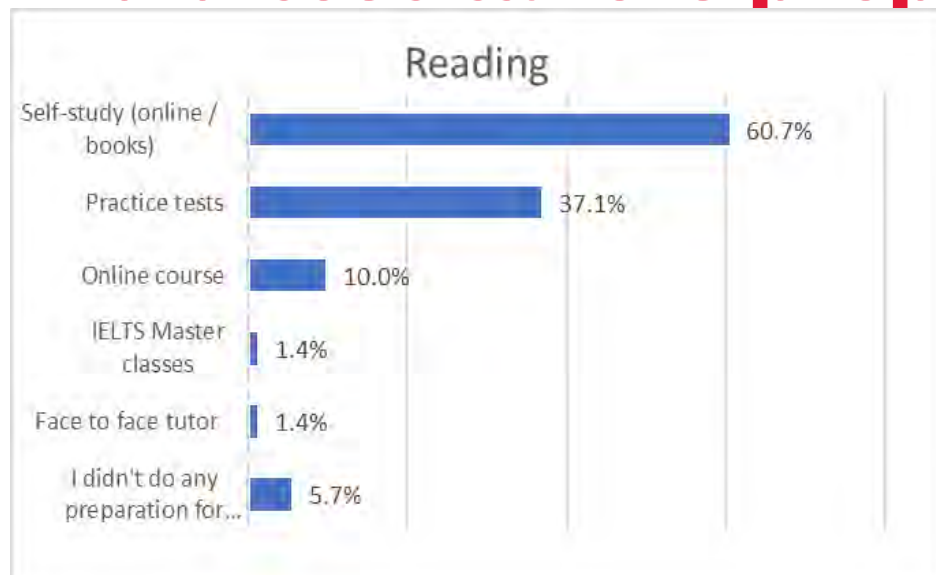
Preparation time

How did your time preparing for IELTS One Skill Retake compare to your time preparing for same skill on the full (4 skills) IELTS test?



$n = 483$

How did test takers prepare for OSR?



Main positive perceptions from questionnaire

| Theme | Frequency |
|---------------------------------|-----------|
| Allows me to focus on one skill | 50 |
| Saves time | 27 |
| Relieves stress | 23 |
| Saves money | 22 |

*"It allows me to **focus on my weakness skill** and not necessarily to redo other parts and **much cheaper.**"*

*"It significantly **alleviates mental stress.**"*

*"OSR gives us the edge of **rectifying those mistakes quickly.**"*

*"When you undertake only one skill test, you have **higher energy and concentration.**"*

Summary

- Generally positive perceptions from test takers
- Main rationale for One Skill Retake: lack of practice, time management, unfamiliarity
- Writing and Reading are most commonly chosen retakes
- 91% of test takers retake the skill with their lowest score
- Score changes:
 - Majority of scores remain the same or increase by up to one band; some may decrease or increase further
 - 45 - 70% of surveyed test takers achieve their desired score (varies by skill)
 - Greater score gain in receptive skills, esp. Listening
 - Change in the Overall band score is, on average, almost identical to that of full test repeaters

Future research plans

- Systematic approach to interviews with receiving organisations (various sectors, countries etc.)
- In-depth analysis of test taker perceptions as to why they did not perform at their best in their original full test
- Extent to which OSR provides an opportunity to mitigate key factors leading to construct irrelevant variance



Thank you!

IELTS

References

- Barkaoui, K. (2017). Examining repeaters' performance on second language proficiency tests: A review and a call for research, *Language Assessment Quarterly*, 14(4), 420-431.
- Green, A., & Van Moere, A. (2020). Repeated test-taking and longitudinal test score analysis. *Language Testing*, 37(4), 475-481.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373-385.
- Hoang, N. T. H., & Hamid, M. O. (2017). 'A fair goal for all?' Australia's language-in-migration policy. *Discourse: Studies in the cultural politics of education*, 38(6), 836-850.
- Kunnan, A. J. (2018). *Evaluating language assessments*. Oxon, UK: Routledge.

References

- Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58, pp. 981-1007. <https://doi.org/10.1111/j.1744-6570.2005.00713.x>
- O'Sullivan, B. & Weir, C.J. (2011) Test development and validation. In B. O'Sullivan (Ed.) *Language Testing: Theories and Practices*. Palgrave. (13-32)
- Roszkowski, M. J., & Spreat, S. (2016). Retaking the SAT may boost scores but this doesn't hurt validity. *Journal of the National College Testing Association*, 2(1), 1-16.
- Streiner, D. L. (2001). Regression toward the mean: Its etiology, diagnosis, and treatment. *Canadian Journal of Psychiatry*, 46, 72-76.
- Walters, F. S. (2022). Ethics and fairness. In G. Fulcher & L. Harding (eds.), *The Routledge handbook of language testing* (pp. 563-577). Oxon, UK: Routledge.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Oxford, UK: Palgrave.

With thanks to:

- Jan Langeslag, British Council
- Ema Bauzyte, Cambridge University Press & Assessment
- Tony Clark, Cambridge University Press & Assessment
- Reza Tasviri, IDP