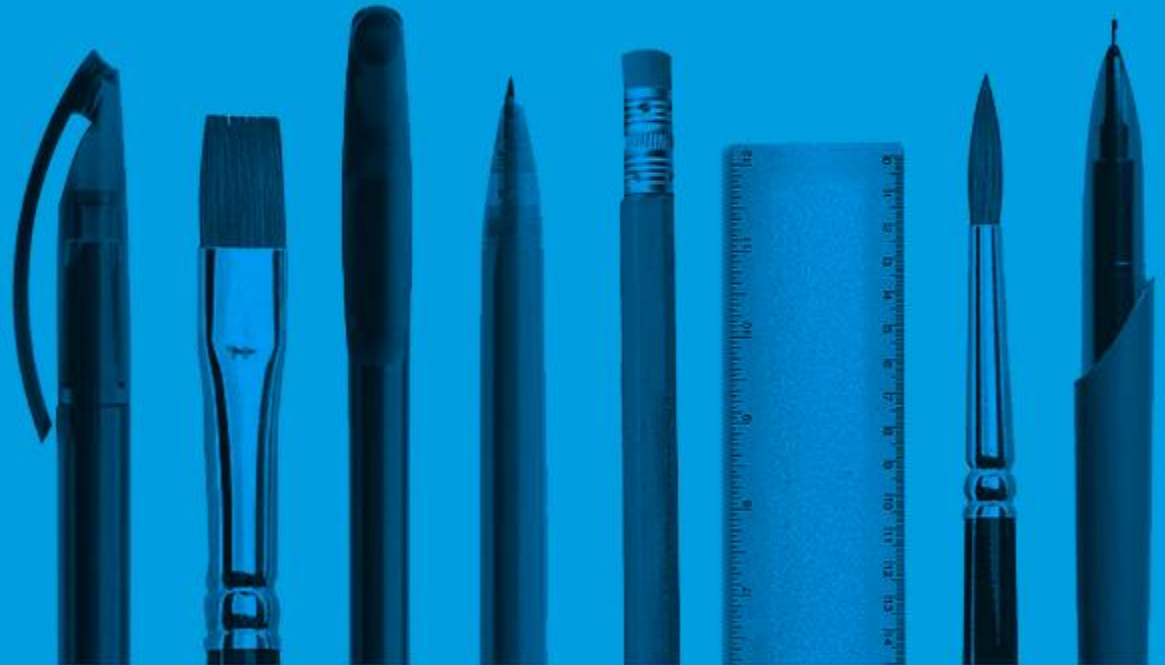


Darogan Hygyrchedd Eitemau wrth Asesu'r Gymraeg

Predicting Item Difficulty in the Assessment of Welsh

Fforwm Asesu Ieithoedd, UKALTA, Prifysgol Abertawe
Language Testing Forum, UKALTA, Swansea University
2019



Emyr Davies
Swyddog Arholiadau Cymraeg i Oedolion
Welsh for Adults Examinations Officer
CBAC-WJEC

Lillie Halton
Rheolwr Dyfarnau a Safonau
Awarding & Standards Manager
CBAC-WJEC

Cydnabyddiaeth

Diolch i:

Joanna Maziarz, Rheolwr Ymchwil (Aseu ac Ansawdd), CBAC.

Alun Evans, Pennaeth Ymchwil a Dadansoddi Ystadegol.

Cyfranogwyr yn y gweithdy.

Acknowledgements

Thanks to:

Joanna Maziarz, Research Manager (Assessment and Quality), WJEC.

Alun Evans, Head of Research and Statistical Analysis.

Workshop participants.

Cyd-destun

Darogan hygyrchedd eitemau mewn profion iaith Gymraeg.

Y profion: Darllen, Gwrando a Gwybodaeth o iaith ar lefelau B1 a B2.

Roedd y cyfranogwyr yn cynnwys arholwyr profiadol, marcwyr, llunwyr eitemau a rhai heb brofiad o'r ymgeiswyr na'r profion.

3 rhan i'r project:

- Rhagweld 160 eitem (holiadur ar-lein)
- Diwrnod gweithdy
- Rhagweld 160 eitem eto (holiadur ar-lein)

Roedd yr eitemau o brofion byw diweddar, felly roedd eu hygyrchedd go iawn yn hysbys.

Context

Predicting difficulty of items in Welsh language tests.

Tests: Reading, Listening and Knowledge of Language at B1 and B2 levels.

Participants included experienced examiners, markers, item-writers and some with no experience of the candidature or the tests.

3 phases to the project:

- Predicting 160 items (on-line survey)
- Workshop day
- Predicting a further 160 items (on-line survey)

Items were from recent live tests, therefore their actual difficulty was known.

Ffactorau hygyrchedd

- Dyma'r ystadegyn a ystyrir amlaf wrth drafod hygyrchedd eitem.
- Fe'i cyfrifir drwy rannu **marc cymedr** yr eitem â'r **marc uchaf** posib.
- Mae'n cynrychioli'r canran o ymgeiswyr a atebodd yr eitem yn gywir.

Facility factors

- Most commonly considered statistic when determining the difficulty of an item.
- Calculated by dividing the **mean mark** of an item by the **maximum mark**.
- Represent the percentage of candidates who answered the item correctly.

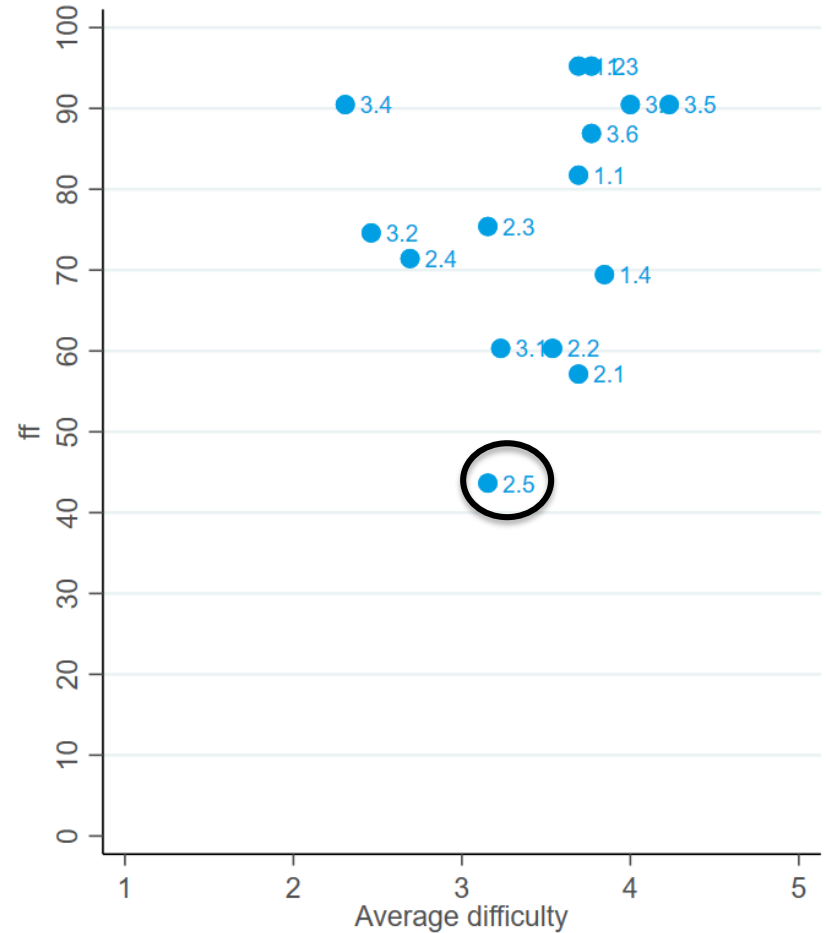
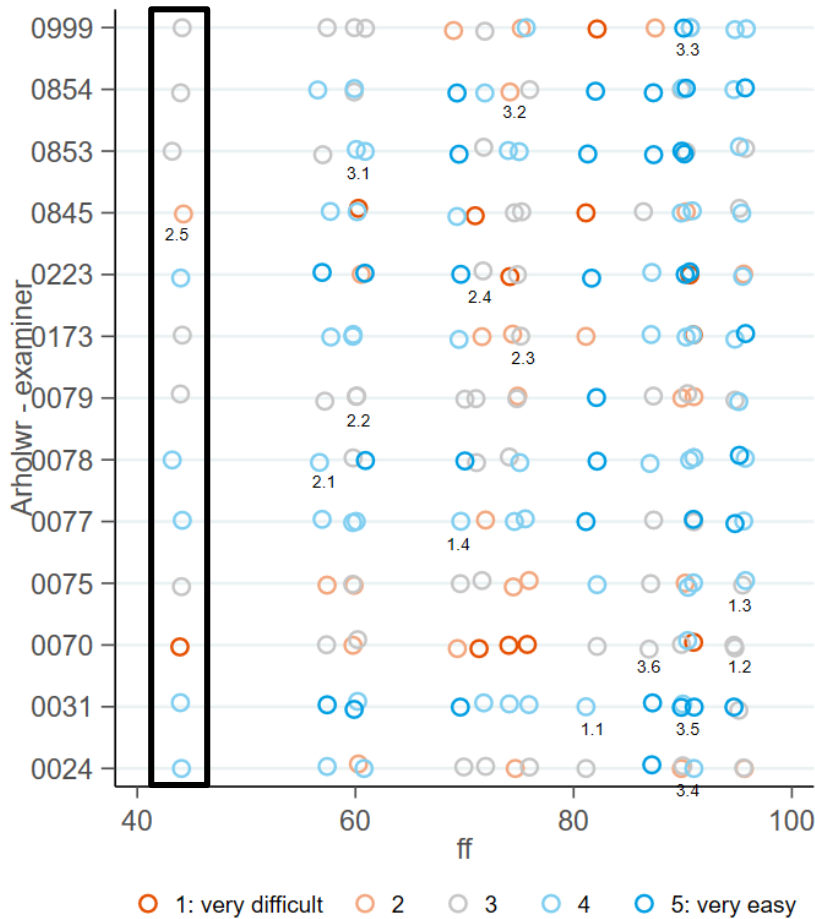
Graddfa Hygyrchedd

1. Anodd iawn – byddai 0-20% o'r ymgeiswyr yn ei gael yn gywir.
2. Eithaf anodd – byddai 20-40% o'r ymgeiswyr yn ei gael yn gywir.
3. Lefel ganolig o ran hygyrchedd – byddai 40-60% yn ei gael yn gywir.
4. Eithaf hawdd byddai 60-80% o'r ymgeiswyr yn ei gael yn gywir.
5. Hawdd iawn – byddai 80-100% o'r ymgeiswyr yn ei gael yn gywir.

Difficulty scale

1. Very difficult – 0-20% of candidates would get this right
2. Quite difficult – 20-40% of candidates would get this right
3. Medium level of difficulty – 40-60% of candidates would get this right
4. Quite easy – 60-80% of candidates would get this right
5. Very easy – 80-100% of candidates would get this right

Uwch Gwrando 2017 - B2 Advanced Listening



Uwch Gwrando 2017 - B2 Advanced Listening



Canlyniadau Results

Ystadegau cydberthyniad ~ Correlation statistics

	Cyn y gweithdy Before workshop		Ar ôl y gweithdy After workshop	
Canolradd Darllen <i>B1 Reading</i>	2015	0.21	2017	0.38
Canolradd Darllen <i>B1 Reading</i>	2016	0.47	2018	0.54
Canolradd Gwrandd <i>B1 Listening</i>	2015	0.87	2017	0.71
Canolradd Gwrandd <i>B1 Listening</i>	2016	0.45	2018	-0.17
Uwch Darllen <i>B2 Reading</i>	2017	0.51	2018	0.20
Uwch Gwrandd <i>B2 Listening</i>	2017	0.34	2018	0.26
Uwch Gwybodaeth o laith B2 <i>Knowledge of Lang.</i>	2017	0.64	2018	0.77

Canolradd Gwrando 2016 - B1 Intermediate Listening



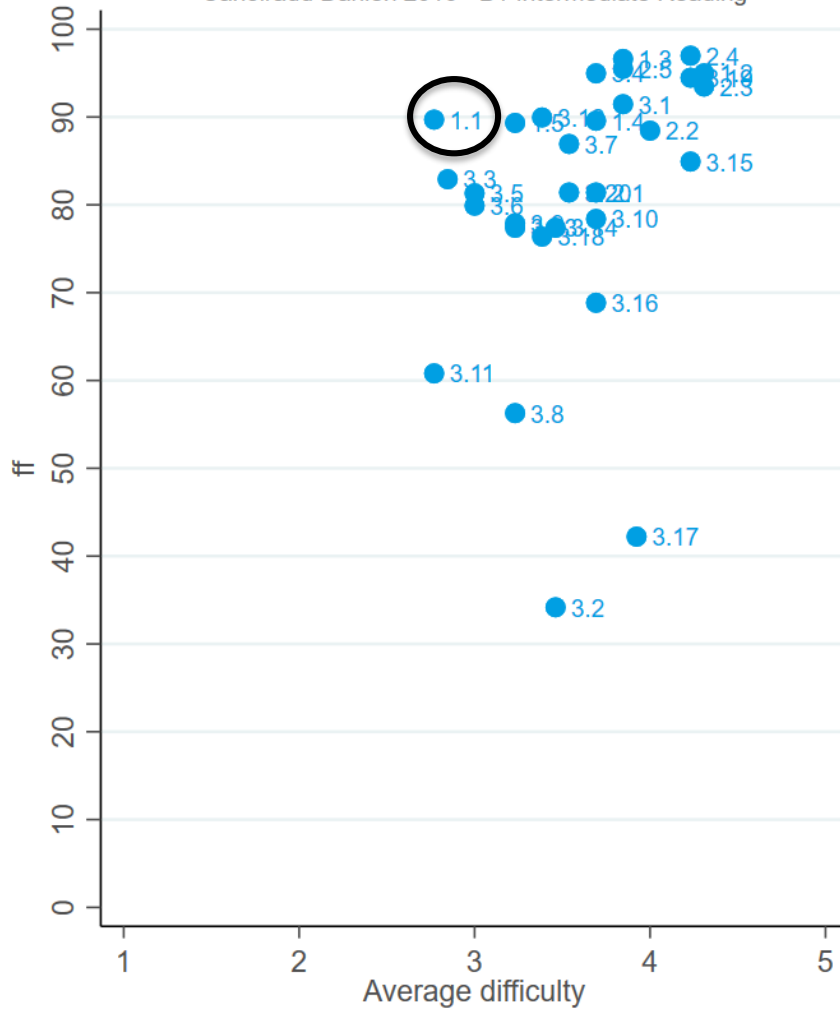
Canolradd Gwrando 2018 - B1 Intermediate Listening



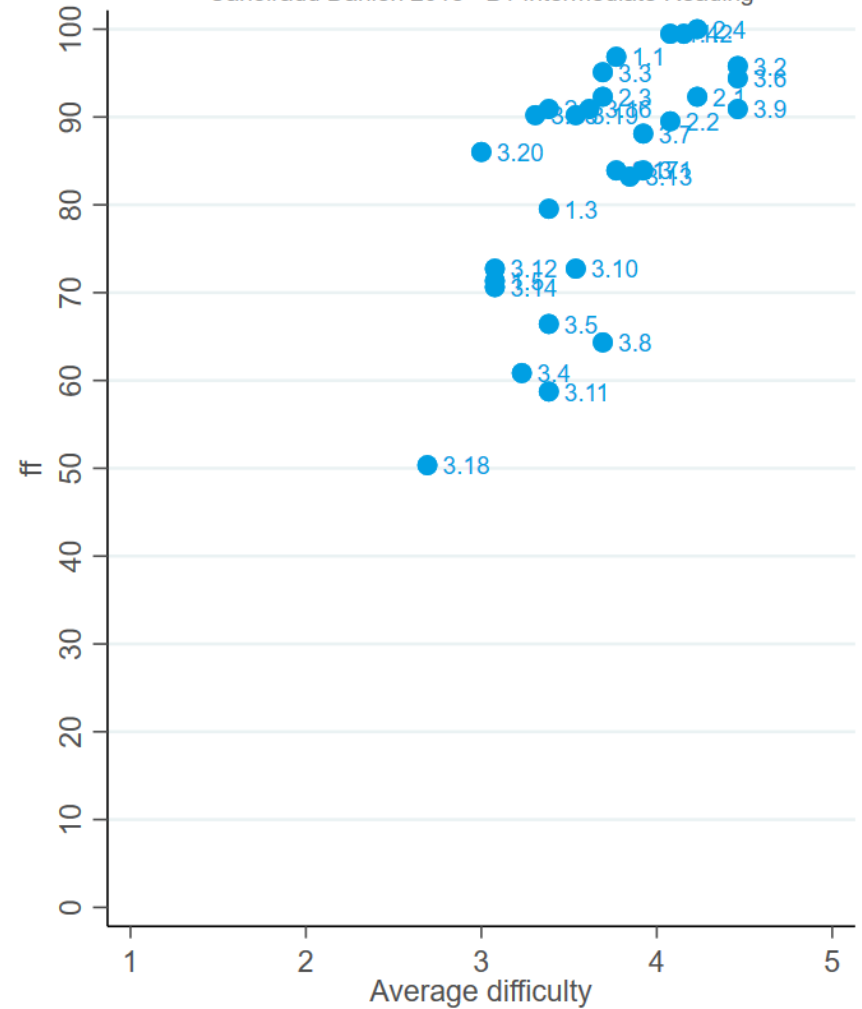
B1 Intermediate 2016 a 2018 Darllen

B1 Intermediate 2016 and 2018 Reading

Canolradd Darllen 2016 - B1 Intermediate Reading



Canolradd Darllen 2018 - B1 Intermediate Reading



Ffactorau sy'n effeithio ar hygyrchedd

e.e.

Geirfa gyfarwydd

Hyd testun, brawddeg a gair

Dwysedd y wybodaeth

Pwnc cyfarwydd

Gwybodaeth amherthnasol

Rhagweladwyedd

Credadwyedd

Cryfder y gwrthdynwyr

Gwybodaeth o ddiwylliant neu'r byd

Gorgyffwrdd geirfaol

Strwythur y testun

Defnydd o ffurfiau llafar/llenyddol

Codgroesi

Effeithiau treigladau

Cyflymder y siarad

Ynganu a goslefu

Amser i brosesu darnau o wybodaeth

Factors affecting accessibility

e.g.

Lexical familiarity

Text, sentence and word length

Information density

Topic familiarity

Redundant information

Predictability

Plausibility

Distracter strength

Cultural or world knowledge

Lexical overlap

Text structure

Use of colloquial/literary forms

Use of code switching

Mutation effects

Speed of delivery

Pronunciation and intonation

Time to process information chunks

‘... identifying text variables which consistently cause difficulty is a complex task. Clearly at some level the syntax and lexis of texts will contribute to [...] difficulty, but the interaction among syntactic, lexical, discourse and topic variables is such that no one variable can be shown to be paramount.’

Alderson, C. 2000 quoted in Weir, C.J. 2005

Casgliadau

Mae'r astudiaeth yn cadarnhau'r dybiaeth fod rhagweld hygyrchedd eitemau'n anodd, ac felly fod cynbrofi'n hanfodol. Allwn ni ddim ymddiried yn unig ym marn arbenigwyr.

Mae hyfforddi'n gwella'r rhagfynegiadau i raddau, ond mae'n her i'r arholwyr mwyaf profiadol hyd yn oed, yn enwedig mewn profion Gwranddo.

Bydd yr astudiaeth yn ein helpu i hyfforddi llunwyr eitemau yn y dyfodol.

Conclusions

The study confirms the view that predicting item accessibility is difficult, and therefore that pretesting is essential. We can not rely solely on expert judgement.

Training improves predictions to some degree, but it remains a challenge even for the most experienced examiners, especially in Listening tests.

The study will help inform how we train item-writers in future.

- Ahmed, A., Pollitt, A. (1999). Curriculum Demands and Question Difficulty. Paper presented at IAEA Conference, Slovenia, May 1999.
- Beinborn, L., Zesch, T., Gurevych, I. (2014) Predicting the Difficulty of Language Proficiency Tests. *Transactions of the Association for Computational Linguistics*, 2, 517-529.
- Buck, G. (2001), *Assessing Listening*, Cambridge Language Assessment Series, CUP.
- Gathercole, V.C., and Thomas, E.M. (2003). Minority Language Survival: Input Factors Influencing the Acquisition of Welsh. *Proceedings of the 4th International Symposium on Bilingualism*.
- Green, A. (2014). *Exploring Language Assessment and Testing. Language in Action*. New York: Routledge
- Johnson, M. (2014). Insights into Contextualised Learning: How do Professional Examiners Construct Shared Understanding Through Feedback. *E-Learning and Digital Media*, 11(4), 363-378.
- Loukina, A., Yoon, S.Y., Sakano, J., Wei, Y., Sheedan, K.M. (2016). Textual complexity as a predictor of difficulty of listening items in language proficiency tests. *Paper presented at Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, p.3245-3253, Osaka, Japan.
- Nettle, D. (2012) Social Scale and Structural Complexity of Human Language. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 367, 1829-1836.

Rush, R.B., Rankin, D.C., White. B.J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education*, 16, 1-10.

Susanti, Y., Nishikawa, H., Tokunaga, T., Hiroyuki, O. (2016) Item Difficulty Analysis of English Vocabulary Questions. *Proceedings of the 8th International Conference on Computer Supported Education (CSEDU 2016)*. 1, 267-274.

Sydorenko, T. (2011). Item Writer Judgments of Item Difficulty Versus Actual Item Difficulty: A Case Study. *Language Assessment Quarterly*, 8, 34-52.

Weir, C.J. (2005) *Language Testing and Validation*, Research and Practice in Applied Linguistics, Palgrave Macmillan.