## LIFE AND WELLBEING SCIENCE

Historical development of item response theory

LIBERATO CAMILLERI

Item response theory (IRT) provides the theoretical framework for measuring latent traits that cannot be observed directly. Examples of latent traits include intelligence, self-efficacy, mental disorder, introversion, emotional difficulty, prosocial behaviour and resilience. A measure for an individual's latent trait can be inferred through a statistical model based on several observable behaviours. For example, an individual's conduct problems (latent variable) can be measured by rating observable behaviours, including fighting, cheating, bullying, lying, stealing, swearing, and losing temper.

An important contribution to IRT literature was made by Lord in 1952 who developed the two-parameter normal ogive model. This was followed by Birnbaum's contribution in 1957 who presented the two-parameter logistic model, which is more mathematically tractable than
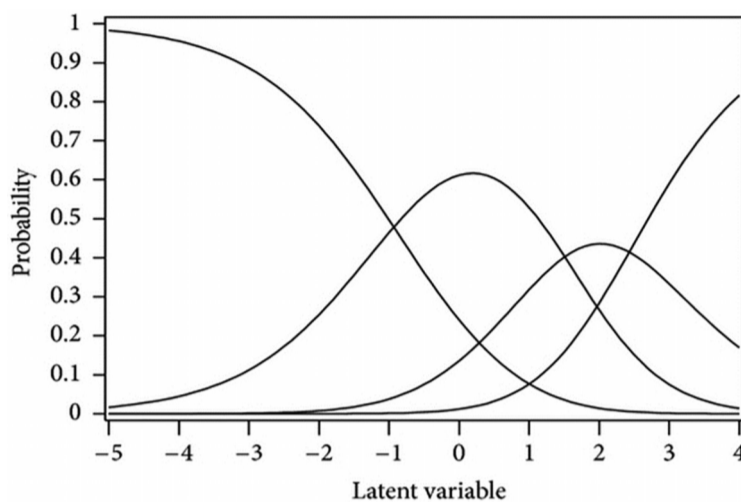
Item characteristic curves for polytomous scaled items.
PHOTO: HTTPS://STATIC-02.HINDAWI.COM/ARTICLES/ISRN/VOLUME-2013/617475/FIGURES/617475.FIG.001.JPG

the normal ogive model. In 1960, Rasch developed the one-parameter and two-parameter Rasch models for binary responses, which enabled the researcher to compare empirical data to assess an instrument's capacity to emulate the properties of fundamental measurement and thus serve as a tool for quantifying unobservable human conditions.

In 1969, Samejima proposed the Graded Response Model for ordinal responses. Samejima also introduced the multi-dimensional IRT model to measure multiple traits in a test. In 1972, Bock presented an IRT model for nominal responses with several categories, which is also appropriate to model responses of multiple-choice items. Bock and Lieberman made several contributions in estimating the IRT model parameters using the marginal maximum likelihood. In 1974, the computer program LOGIST was developed to estimate IRT model parameters. In 1978, Andrich developed the Rating Scale Model (RSM) for

polytomous responses, which is a generalisation of the Rasch model for dichotomous responses. In 1982, Masters proposed the Partial Credit Model (PCM) for ordinal responses and in 1991, Muraki formulated the Generalised Partial Credit Model (GPCM), relaxing the assumption of uniform discriminating power of test items.

In 1992, Wainer developed a computer adaptive testing method with the aim of reducing the testing time and, in 1993, Holland and Wainer introduced differential item functioning (DIF) in IRT models. Moreover, Gibbons developed further the full information bi-factor model in order to consider polytomous item responses.

In 1996, Zimowski, Muraki and Bock developed an IRT model for multiple groups. These models were essential for researchers to examine the test items in distinct groups. In the last two decades, most of the advances in IRT were linked to the development of computer programs, including Itm (2006), IRTPRO (2011), mirt (2012) and ACER ConQuest (2015), which all facilitated the estimation of IRT model parameters.

Liberato Camilleri is statistics professor, University of Malta.

DID YOU KNOW?

These points were inferred from local studies using IRT models:
- Maltese males tend to be more assertive, risk-prone, thrill-seeking, tough-minded and have higher self-esteem than females. In contrast, females, on average, tend to be more prosocial, sociable, sensitive, and have higher extraversion and self-efficacy than males.
- Maltese females tend to have higher levels of anxiety, stress and panic attacks than males. In contrast, males, on average, are more likely to be diagnosed with attention deficit disorder (ADD) and tend to have more hyperactivity and conduct problems than females.
- Maltese male students engage in bullying more than females and this applies to all types of bullying. However, gender differences in verbal and physical bullying are more conspicuous than relationship bullying.
- In Maltese primary schools, boys tend to perform better than girls in mathematics and science. However, this trend is reversed in secondary schools.
For more trivia see: www.um.edu.mt/think.

SOUND BITES

- **From the international scene:** Several large-scale educational survey assessments such as the Programme for International Student Assessment (PISA) involve the use of item response theory (IRT) models. The purpose of these models is to calibrate and evaluate test items, questionnaires and other instruments, and to score subjects on their abilities, attitudes, behaviours and other latent traits.
- **From the local scene:** The author of this page (Liberato Camilleri) applied IRT models to two studies. The first showed that Maltese xenophobic attitude towards irregular immigrants originates from concerns that these immigrants would increase crime rates and hamper our culture/values, rather than concerns related to job-opportunity reduction. The second showed that Maltese sentiment in favour of divorce is more likely to occur in the presence of domestic violence and adultery rather than lengthy illness, financial problems and inability to have children.

For more science news, listen to Radio Mocha on www.fb.com/RadioMochaMalta/.

PHOTO OF THE WEEK
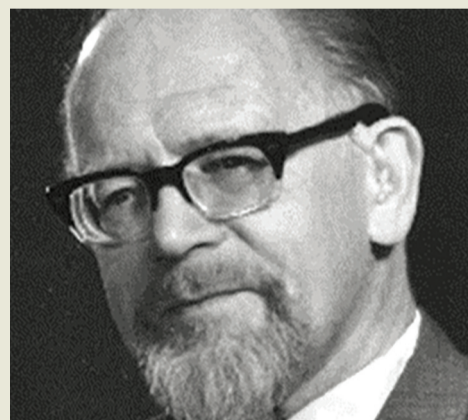
Georg Rasch (1901-1980) was a Danish statistician. He commenced his studies in mathematics at the University of Copenhagen in 1919, and completed his master's degree in 1925 and his PhD in 1930. In 1935, he joined the Danish Bio-Statistical Department as a statistical consultant. In 1960, he published his work on the Rasch model, which is still widely applied in statistics. In 1961, he was appointed a professor of statistics at the University of Copenhagen. He was knighted in 1967. PHOTO: WWW. RASCH.ORG/GIFS/RASCH.GIF

There are several misconceptions in Maltese education. Through appropriate statistical analysis on local data, it is possible to confirm facts and debunk myths.
- Resilient students are defined as high academic achievers despite being socio-economically disadvantaged. A common perception is that resilient students are more prevalent in Church and Independent schools. This may be true for male students; however, female resilient students are more prevalent in state schools.
- Another perception is that bullying is more prevalent in state schools than Independent and Church schools. The PISA 2018 Malta report shows that bullying is prevalent in all school types. Physical bullying (25.3%) is the most common, followed by verbal bullying (19.9%), relationship bullying (15.5%), and cyber bullying (15.3%).
- It is often thought that academic performance is unrelated to the child's socio-economic status. However, several international studies show that attainment in mathematics, science and reading comprehension is positively related to socio-economic status. Such children have more academic resources at home and tend to receive more parental support.
- The gender stereotyping in students' career expectation is diminishing. The PISA 2018 Malta Report reveals that the most cited occupations by girls are similar to those cited by boys.