Item Response Theory on Ergonomics and Human Factors Researches

Jonhatan Magno Norte da Silva¹, Antonio Cezar Bornia², Leila Amaral Gontijo², Luiz Bueno da Silva³

¹Department of Production Engineering, Federal University of Santa Catarina, Santa Catarina, Brazil; ²Department of Production Engineering, Federal University of Alagoas, Alagoas, Brazil; ³Department of Production Engineering, Federal University of Paraíba, Paraíba, Brazil

DESCRIPTION

It is common practice in ergonomic studies to use instruments that seek to measure, through items or questions, some perception, sensation, skill or characteristic (latent trait) of individuals or elements that are part of the workplace. The use of these instruments is preferable, especially when direct medication of the latent trait is difficult, complicated or expensive. Examples of latent traits in ergonomics are the ‘degree of adequacy of students with school desks’ [1], the ‘comfort of aircraft seats’ perceived by users [2], and the ‘musculoskeletal discomfort experienced by workers’ after the workday [3,4].

In psychometry, two theories have been used to quantify latent traits. The first one is the Classic Test Theory (CTT), where the direct or weighted sum of the items generates a score to represent quantitatively the latent trait. The second is Item Response Theory (IRT), in which response patterns are analyzed, so that different response patterns lead to different scores for the latent trait assessed [5]. Thus, the IRT is not based on the number of responses marked positively or negatively to a given set of items, but on estimating the probability that an individual will mark a given response when he has a certain skill, perception or characteristic [6].

Thus, there are differences between CTT and IRT. Critics claim that in CTT the differences between items in an Instrument are ignored leading to skewed results [7]. Thus, scores and metrics generated by IRT are superior in terms of advantages to those generated by CTT, as they present less bias and better adjustment of data in general [8].

Embretson et al. [9] state that in CTT it is not possible to ensure that parameters of individuals are independent of the items presented to them, something guaranteed by the IRT. In practical terms, when estimating an individual’s latent trait via CTT it is likely that a different result for this same latent trait would be obtained if items of the Instrument were modified (removed or inserted). This is because in CTT the conclusions about the latent trait depend on the test or questionnaire as a whole, and not on each item that composes it [10]. In ergonomics, this fact causes some uncertainty in the decision-making process, since in most cases ergonomic interventions are guided by scores of tools, instruments, questionnaires and diagrams composed by items.

However, some factors hinder and inhibit the use of IRT in ergonomic studies. The main one is the need for large samples for a better calibration of the item parameters and generation of the respondents’ scores [11]. However, [9] point out that, through the IRT, smaller samples can also be worked on, provided that this sample has in its composition individuals with different latent traits and, therefore, that they answer the Instrument items differently, generating some response patterns. Evidently, in larger samples, individuals with different latent traits will be more easily found, which makes researchers in ergonomics choose to use CTT in most cases.

Going beyond estimating the respondents’ scores, IRT allows the creation of metrics that position the items and the individuals’ ability on the same scale, facilitating the interpretation of the different latent trait values. In CTT, as a rule, the average or median scores of individuals are calculated as a parameter, and those individuals with a score value above these parameters are classified as high exposure or skill (depending on the latent trait analyzed), and vice versa for those with score values below mean or median [12]. In the IRT, the anchoring method, like that proposed by [13] or by [14], create different levels or classifications for latent traits. Therefore, individuals, depending on the value of their latent trait, will be positioned at some level of the scale, and will have their latent trait classified.

The studies by [14] and [3] used the anchoring proposed by [13] to generate their scales. [15] generated six levels of discomfort (from Without discomfort to Maximum discomfort) for aircraft seating [3] generated a metric of four levels of musculoskeletal discomfort (from Mild discomfort to Maximum discomfort) based on the symptoms reported by workers. Using the [14], the research by [4] classified the musculoskeletal discomfort in the workers’ upper body into six levels (from Minimal discomfort to Maximum discomfort). Through these procedures and interventions of an ergonomic nature they can be better targeted and effective, since they are supported by metrics with less error and bias.

Through the study by [13] it was possible to observe that the
aspects associated with the hardness of the seats and noise present during the trip are those that most contribute to the discomfort of those who perform domestic flights, followed by factors associated with body movement and tension and the presence of neighbors occupying the side seat (fighting for armrests). Therefore, it is clear which variables should be addressed as a priority in order to improve the flight experience of aircraft users. The findings by [4] showed, based on the response patterns generated by the IRT, that the right and left elbows, in addition to the cervical, trapezoidal, middle and upper back, are the regions marked exclusively by individuals with higher degrees of musculoskeletal discomfort. In the same study, it was evidenced that the response patterns framed individuals with different degrees of discomfort, facilitating the screening process of these individuals. Such screening can be useful in the management of musculoskeletal disorders, which is a serious health problem in several countries [10-16]. Therefore, depending on the study’s objectives in ergonomics, valuable findings can be extracted from the scales built based on IRT.

However, caution is necessary, as the purpose of this short comment is not to state that studies based on the assumptions of CTT are wrong in their conclusions, but rather to highlight that the limitations and weaknesses of many studies in ergonomics can be overcome with the use of IRT, bringing more accurate and reliable results for ergonomists to make better supported decisions. It is also expected to encourage the use of IRT in ergonomic studies as there is a lack of research on this topic, mainly because the vast majority of instruments in ergonomics have not been revalidated by the IRT perspective [17], generating relevant scientific gaps.

REFERENCES