Evaluation of normative data of the instrument European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) in Brazilian adults

Avaliação dos dados normativos do instrumento European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) em adultos brasileiros

Evaluación de datos normativos del instrumento del instrumento European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) en adultos brasileños

Fábio Luiz Mialhe¹, Flávio Rebustini².

ABSTRACT

Objective: To investigate the normative data of the Brazilian Portuguese version of the instrument for measuring health literacy European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) in Brazilian adults. Methods: This is a methodological study in which the normative data of scores of HLS-EU-Q6 was evaluated in a sample of 783 adult users of the Brazilian public health system (SUS). The normative analysis went through three stages: descriptive analysis of the items and the instrument scores; analysis of the distribution of the score and confirmation that the cutoffs established in the distribution were able to effectively identify the participants in each range. Results: It was observed that all instrument items had a median of 3 and the median of the total score was 17 points. The results suggest that a model with 3 cutting lines and with different scores two proposals by the authors of the original study turned the classification capacity of the instrument more accurate for this population. Conclusion: It is concluded that the classification criteria of the HLS-EU-Q6 to be used in Brazilian adults should be different from those proposed by the authors of the original study, so as not to generate biases in the measurement of health literacy levels in this population.

Keywords: Health literacy, Validation studies, Unified health system.

RESUMO

Objetivo: Investigar os dados de normatização da versão em português do Brasil do instrumento para aferição do letramento em saúde European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) em adultos brasileiros. Métodos: Trata-se de um estudo metodológico no qual se avaliou a normatização dos escores do HLS-EU-Q6 em uma amostra de 783 adultos usuários do sistema público de saúde brasileiro (SUS). As análises de normatização passaram por três etapas: análise descritiva dos itens e do escore do instrumento; análise de distribuição dos escores e a confirmação de que os cortes estabelecidos na distribuição conseguiram efetivamente identificar os participantes em cada faixa. Resultados: Observou-se que todos os itens do instrumento apresentaram a mediana 3 e a mediana do escore total foi de 17 pontos. Os resultados apontaram que um modelo com 3 faixas de corte e com escores diferentes dos propostos pelos autores do estudo original tornou a capacidade de classificação do instrumento mais precisa para essa população.

Keywords: Leitura e saúde, Validação de estudos, Sistema de saúde.

1 Universidade Estadual de Campinas (UNICAMP), Faculdade de Odontologia de Piracicaba, Piracicaba - SP.
2 Universidade de São Paulo (USP), Escola de Artes, Ciências e Humanidades, São Paulo - SP.
população. **Conclusão:** Conclui-se que os critérios de classificação do HLS-EU-Q6 a serem utilizados em adultos brasileiros devem ser diferentes dos propostos pelos autores do estudo original, de forma a não gerar vieses de aferição nos níveis de letramento em saúde nessa população.

**Palavras-chave:** Letramento em saúde, Estudos de validação, Sistema único de saúde.

**RESUMEN**

**Objetivo:** Investigar los datos de estandarización de la versión en portugués brasileño del instrumento para medir la alfabetización en salud European Health Literacy Survey Questionnaire short-short form (HLS-EU-Q6) en adultos brasileños. **Métodos:** Se trata de un estudio metodológico en el que se evaluó la estandarización de los puntajes HLS-EU-Q6 en una muestra de 783 usuarios adultos del sistema público de salud (SUS) brasileño. El análisis de estandarización pasó por tres etapas: análisis descriptivo de los ítems y puntuación del instrumento; análisis de la distribución de la puntuación y la confirmación de que los puntos de corte establecidos en la distribución lograron identificar efectivamente a los participantes en cada rango. **Resultados:** Se observó que todos los ítems del instrumento tuvieron una mediana de 3 y la mediana de la puntuación total fue de 17 puntos. Los resultados mostraron que un modelo con 3 rangos de corte y con puntajes diferentes a los propuestos por los autores del estudio original hizo más precisa la capacidad de clasificación del instrumento para esta población. **Conclusion:** Se concluye que los criterios de clasificación HLS-EU-Q6 a ser utilizados en adultos brasileños deben ser diferentes a los propuestos por los autores del estudio original, para no generar sesgos en la medición de los niveles de alfabetización en salud en esta población.

**Palabras clave:** Alfabetización en salud, Estudio de validación, Sistema único de salud.

**INTRODUCCIÓN**

La alfabetización en salud (HL) es considerada por la Organización Mundial de la Salud como un importante determinante de la salud y ha sido estudiada con mayor interés por varios autores en los últimos decenios (KICKBUSCH I, et al., 2013; LIU C, et al., 2020). La HL se define como el conocimiento, motivación y habilidades de las personas para acceder, entender, evaluar y aplicar la información en salud, de forma que puedan tomar decisiones y realizar acciones diarias en relación con la atención de la salud, la prevención de enfermedades y la promoción de la salud para mantener o mejorar la calidad de vida (SØRENSEN K, et al., 2013).

Según el Glosario de Términos de la Promoción de la Salud, la HL “representa el conocimiento personal y competencias que se acumulan a lo largo de las actividades diarias, las interacciones sociales y entre generaciones. El conocimiento personal y competencias son mediados por las estructuras organizacionales y la disponibilidad de recursos que permiten a las personas acceder, entender, evaluar y usar la información en formas que promuevan y mantengan la buena salud y el bienestar por sí mismos y los de ellos” (ORGANIZACIÓN MUNDIAL DE LA SALUD, 2021).

Diversos instrumentos se han desarrollado hasta ahora para medir los niveles de HL en la población, sin embargo, muchos de ellos solo evalúan aspectos de la alfabetización funcional en salud, caracterizados por la capacidad de leer y entender la información relacionada con la salud (SØRENSEN K, et al., 2013; SØRENSEN K, et al., 2015; TAVOUSI M, et al., 2022). Sin embargo, HL “significa más que la habilidad de tener acceso a sitios web, leer folletos y seguir las pautas prescritas de búsqueda de salud. Incluye la habilidad para ejercer un juicio crítico acerca de la información y los recursos de salud, así como la habilidad para interactuar y expresar los necesidades personales y sociales para promover la salud” (ORGANIZACIÓN MUNDIAL DE LA SALUD, 2021). Por lo tanto, estos instrumentos no son suficientes para medir de manera más comprehensiva el constructo de HL.

Entre los instrumentos desarrollados para medir HL en un contexto multidimensional, el European Health Literacy Survey Questionnaire (HLS-EU-Q47) evalúa la dificultad percibida por las personas al realizar ciertas actividades relacionadas con la salud (KICKBUSCH I , et al., 2013; SØRENSEN K, et al., 2013; SØRENSEN K, et al., 2015). Es uno de los instrumentos más utilizados a nivel mundial y ha sido adaptado a varios idiomas (TAVOUSI M, et al., 2022). Otros, como el Health Literacy Questionnaire (HLQ)
developed in Australia demonstrated good psychometric properties and have been adapted to the Brazilian Portuguese language (OSBORNE RH, et al., 2013; MORAES KL, et al., 2021).

The HLS-EU-Q47 instrument is composed of 47 questions, which assess the self-reported difficulties of individuals to make decisions in three contexts: health care, disease prevention and health promotion. Despite its qualities to measure the HL construct more comprehensively, it is considered too long for the purposes of individual assessments in the daily context of health services (PELIKAN JM and GANAHL K, 2017).

Therefore, two short versions were developed from it, one with 16 questions (HLS-EU-Q16), named short-form and other with only 6 questions named short-short-form (HLS-EU-Q6) (PELIKAN JM and GANAHL K, 2017). They were developed by a team of the European Consortium using the HLS-EU-Q47 data based on Item Response Theory and Rasch Analysis (PELIKAN JM, et al., 2014). The HLS-EU-Q6 instrument presents 6 of the 16 items of HLS-EU-Q16 and takes about a minute of interviewing time (PELIKAN JM, et al., 2014).

Very few studies evaluated the characteristics of HSL-EU-Q6 indicate that the instrument present good psychometric properties (LORINI C, et al., 2019; ROUQUETTE A, et al., 2018). In relation to the scoring process, the developers of the instrument recommended three levels for the scale classification: inadequate, problematic and sufficient HL and stated that the “criterion was correct classification, as far as possible, compared to Q47 levels (PELIKAN JM, et al., 2014).

Therefore, no studies to date have assessed the validity of this form of categorization suggested by the authors through using a more robust analyzes to establish valid and reliable cutoff criteria for classifying HL levels using HLS-EU-Q6. If individuals are classified incorrectly on their HL level, this fact can have important implications for the planning of health service activities.

HLS-EU-Q6 was previously adapted to Brazilian Portuguese language and their psychometric properties evaluated in Brazilian adults (MIALHE FL, et al., 2021). However, the validity of the classification scores for Brazilian adults was not evaluated until the development of this research. Therefore, the aim of this study was to investigate the normative data of the Brazilian Portuguese version of HLS-EU-Q6 in Brazilian adults.

METHODS

The evaluation of the psychometric properties of HLS-EU-Q6 was carried out through a cross-sectional study with a convenience sample of 783 Brazilian adults. The detailed methodology was also described in a previous publication (MIALHE FL, et al., 2021). In short, 320 participants were from the city of São Paulo/SP and 293 from the city of Piracicaba/SP, both living next to Primary Health Units. The households were randomly selected, and adults who were present at the time of the visit were invited to participate in the research. In addition, 50 patients from a reference institution in cancer treatment in Fortaleza/CE and 120 individuals living next to a Primary Health Unit from Aparecida de Goiânia/GO, Brazil were invited to participate in the research.

The HLS-EU-Q6 consists of six questions from HLS-EU-Q47 with five response categories as follows: very difficult=1, difficult=2, easy=3, very easy=4, and a fifth alternative for when participants did not answer or did not have a definitive answer. This is labelled as “don’t know/refused and is treated as missing value (value 0). The instrument present the following statement: [On a scale that goes from “very easy” to “very difficult”, how easily you can] 1. assess when you need a second opinion from another doctor? 2. use the information that your doctor gives you to make decisions about your illness?; 3. find information on how to deal with mental health problems, such as stress or depression?; 4. Assess whether the information on health risks available in the media is reliable? (e.g. TV, Internet or other means of communication); 5. find information about activities that are good for your mental well-being? (e.g. meditation, exercise, walking, pilates, etc.); 6. understand the information available in the media on how to stay healthier? (e.g. Internet, newspapers, magazines)].

According to the authors who developed the instrument, the final individual score is a mean calculated by summing up the answers to the six questions divided by the number of items answered (PELIKAN JM, et al.,
The score is calculated as long as at least five of the six questions are answered differently from 0, and varies between 1 and 4, with higher values indicating better levels of HL.

According to them, the final score values classify individuals according to three levels of HL: (likely) inadequate (≤ 2); (likely) problematic (> 2 and ≤ 3); and (likely) sufficient (> 3) (PELIKAN JM, et al., 2014). Moreover, another General HL index (G-HL16 index) can be calculated as follows: G-HL16 index = (mean-1)x(50/3). Then, four levels of HL were defined: inadequate (0-25), problematic (25.1-33), sufficient (33.1-42) and excellent (42.1-50) (PEDRO AR, et al., 2023).

The study of standardization goes through three stages: the first is the descriptive analysis of the items and the instrument score; second, the analysis of the distribution of the score and, finally, the confirmation that the cuts established in the distribution are able to effectively identify the participants in each range. In this way, a descriptive exploratory study of the general scores of the HLS-EU-Q6 was initially carried out. The results of the items and the total score were represented by the frequency of responses, median (Md), interquartile range (IQR), range (amp), minimum (min), maximum (max).

As a way of identifying the instrument items that can serve as relevant markers and their weights for the assessment of the NCSM, a regularized linear regression with elastic net was applied, adopting the instrument score as the dependent variable and the items as predictors. Holdout was applied as a cross-validation technique with 50% of the sample for training and 50% for testing. The first stage of normalization was carried out with the identification of the score cuts based on the distribution of participants. Despite this process, since the distribution of participants is recurrent in standardization studies, it can be distorted, as the score is not directly analyzed, but considered as a consequence of the position of the participants in the cutoff points.

In order to obtain greater precision regarding the cuts in the ranges and to verify the predictive capacity of classifying the individuals, a discriminant analysis was used for each of the ranges and the scores of the HLS-EU-Q6. Discriminant analysis aims to understand group differences and predict the probability that an entity (individual or object) belongs to a specific class or group based on several independent metric variables (HAIR JR, et al., 2019) allowing to determine which of the independent variables are most responsible for the differences in the average score profiles of the two or more groups (HAIR JR, et al., 2019).

Boedeker P and Kearns NT (2019) point out that discriminant analysis has proven to perform better than other techniques, such as: logistic regression, multinomial regression, Random forest, k-nearest neighbor and Support-vector machines, when the objective is to classify a member of a group. In this way, it can be used to confirm whether the cutoffs established by the distribution have the property of correctly classifying individuals into the bands. Data were analyzed using the statistical program SPSS v.23 and JASP 16.04. The research project was approved by the Research Ethics Committee (CAAE: 58131216.5.0000.5418; technical opinion nº 2.786.634).

RESULTS

This study included 783 individuals whose mean age was 38.6 years, and 68.1% were female. In addition, 33.5% had studied up to elementary school. Table 1 presents the descriptive data for the items and for the instrument score. An important point is that the amplitude of the scale was used, which indicates, preliminarily, that it is effective in capturing nuances of the participants’ responses. All items had the median 3.0 as the central point, with items 2, 5 and 6 having an IIQ (interquartile range) = 0 and items 1, 3 and 4 having an IIQ = 1. The median of the total score was 17 points (IIQ = 2) with a range ranging from 6 to 24, so the sample presented scores in all possible ranges.
Table 1 - Descriptive Statistics of Items and response frequency on the scale.

<table>
<thead>
<tr>
<th>Measures of central tendency and dispersion</th>
<th>Item Response Frequency (N/%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. assess when you need a second opinion from another doctor?</td>
<td>3.0</td>
</tr>
<tr>
<td>Q2. use the information that your doctor gives you to make decisions about your illness?</td>
<td>3.0</td>
</tr>
<tr>
<td>Q3. find information on how to deal with mental health problems, such as stress or depression?</td>
<td>3.0</td>
</tr>
<tr>
<td>Q4. assess whether the information on health risks available in the media is reliable?</td>
<td>3.0</td>
</tr>
<tr>
<td>Q5. find information about activities that are good for your mental well-being?</td>
<td>3.0</td>
</tr>
<tr>
<td>Q6. understand the information available in the media on how to stay healthier?</td>
<td>3.0</td>
</tr>
<tr>
<td><strong>Total Score</strong></td>
<td>17.0</td>
</tr>
</tbody>
</table>

Legend: Md – median; IIQ – interquartile range; Amp – amplitude; Min – minimum; Max – maximum; SD – standard deviation.


Due to these score results, it was decided to test 3 types of ranges: the first classification model with 2 ranges was composed of the first cut in the median, thus, from 6 to 16 points (insufficient) and 17 to 24 points (sufficient). The second model was tested with 3 ranges: low (6 to 15 points), moderate (16 to 17 points) and high (from 18 to 24).

The third model proposed would be by quartiles; however, the second and third quartiles would have only one score, 16 for the 2nd and 17 for the 3rd quartile, which makes it unproductive and does not improve the accuracy of classification and interpretation of data.

The results of the discriminant analysis for model 1 were: MBox = 4.87; p < 0.001; λwilks = 0.43; F₁(1, 781) = 1033.23; p < 0.001; canonical correlation = 0.75, and it was able to correctly classify 100.00% of the original cases. Model 2 presented a result for the discriminant analysis of MBox = 319.76; p < 0.001; λwilks = 0.29; F₂(2, 780) = 946.66; p < 0.001; canonical correlation = 0.84. Model 2 was also able to correctly classify 100% of the cases. Model 3 presented a result for the discriminant analysis of MBox = 6.84; p < 0.009; λwilks = 0.28; F₃(3, 779) = 666.58; p < 0.001; canonical correlation = 0.84 and was also able to correctly classify 64.6% of the cases.

As both classification of models 1 (with 2 ranges) and 2 (with 3 ranges) were able to correctly classify the participants, it is recommended the use of the model 2 with 3 ranges, as shown in Table 2, because HL is not a dichotomous construct (high and low), but has different levels, as already demonstrated in the literature (TAVOUSI M, et al., 2022).
Table 2 - Classification and interpretation of HLS 6 scores in Brazilian adult sample.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Percentile range</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low/inadequate</td>
<td>Up to 25</td>
<td>6 to 15 points</td>
</tr>
<tr>
<td>Moderate/problematic</td>
<td>26 to 75</td>
<td>16 to 17 points</td>
</tr>
<tr>
<td>High/sufficient</td>
<td>76 to 100</td>
<td>18 to 24 points</td>
</tr>
</tbody>
</table>

**Source:** Mialhe FL and Rebustini F, 2023.

According to the classification criteria presented above, 23.5% of the sample presented high/sufficient HL levels, 30.5% moderate/problematic HL levels and 46.0% present low/inadequate HL levels.

Once the classification and interpretation ranges had been established, it was essential to understand the impact of the items and their influence on the formation of the score. For this, a boosting regression (machine learning technique) with Gaussian function loss and cross-validation with 5-Kfold using the total score of the instrument and the items as predictors. The regression had a Shrinkage = 0.1 with an MSE (mean standard error) in the training bank of 0.020 and 0.017 in the test bank. The model had an $R^2 = 98.4\%$ with a Mean Absolute Percentage Error (MAPE) = 24.17\% and an RMSE (root mean squared error) = 0.13.

Table 3 indicates the relative influence of the instrument items. The values indicate items 3 and 6 have highest influence in the score. It is interesting to point out that both deal with issues of obtaining information regarding health problems.

Table 3 – Relative influence of item in the score.

<table>
<thead>
<tr>
<th>Item</th>
<th>Relative Influence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. assess when you need a second opinion from another doctor?</td>
<td>17.02</td>
</tr>
<tr>
<td>Q2. use the information that your doctor gives you to make decisions about your illness?</td>
<td>9.58</td>
</tr>
<tr>
<td>Q3. find information on how to deal with mental health problems, such as stress or depression?</td>
<td>22.14</td>
</tr>
<tr>
<td>Q4. assess whether the information on health risks available in the media is reliable?</td>
<td>15.51</td>
</tr>
<tr>
<td>Q5. find information about activities that are good for your mental well-being?</td>
<td>14.26</td>
</tr>
<tr>
<td>Q6. understand the information available in the media on how to stay healthier?</td>
<td>21.45</td>
</tr>
</tbody>
</table>

**Source:** Mialhe FL and Rebustini F, 2023.

DISCUSSION

The present study demonstrated through a rigorous standardization process, the determination of the best cutoff points for classifying the scores of the HLS-EU-Q6 instrument in Brazilian adults.

To date, this is the first study that evaluated the validity of the scores of HLS-EU-Q6 proposed by the authors who developed the instrument. All other studies using that instrument only replicated the scores proposed by original authors to classify the HL levels of individuals in their samples (PELIKAN JM, et al., 2014; AMOAH PA, et al., 2017; VANDENBOSCH J, et al., 2018; SCHINCKUS L, et al., 2018; LORINI C, et al., 2019; ROUQUETTE A, et al., 2018).

Our results indicated that the way of scoring the instrument in Brazilian adults should be substantially different from that proposed by the original authors in order not to generate incorrect interpretations. In the original study, the HLS-EU-Q6 items were selected from HLS-EU-Q16 items based on higher item difficult and the authors recommended that the score should be calculated as a mean score by the sum of answers divided by the number of items answered (PELIKAN JM, et al., 2014). Thus, scores range from 1 to 4. According to the authors, three levels for the scale were defined, that is, “likely” inadequate (≤ 2); “likely” problematic (> 2 and ≤ 3); and “likely” sufficient (> 3) (PELIKAN JM, et al., 2014).
In the scoring proposal for the Brazilian sample, the division of the sum of responses by the number of items answered was not used. However, if that were the case, the classification criteria would be: low or “inadequate” (≤ 2); moderate or “problematic” (2.17-2.33) and high or “sufficient” (2.5-4.0). Therefore, the scale of normed scores of HLS-EU-Q6 in Brazilian adults was very different from that recommended by the original authors and used in European and other countries (PELIKAN JM, et al., 2014; AMOAH PA, et al., 2017; VANDENBOSCH J, et al., 2018; SCHINCKUS L, et al., 2018; LORINI C, et al., 2019; ROUQUETTE A, et al., 2018).

This fact has important impacts on the classification of individuals according to their HL levels. According to the classification proposed by the original authors of the HLS-EU-Q6, only 2% of the participants of this study would be classified as having sufficient levels of HL; 51.7% with problematic levels, and 46.3% with inadequate levels.

However, according to the classification proposed in the present study, 23.5% of the sample presents high/sufficient HL levels, 30.5% moderate/problematic HL levels and 46.0% presents low/inadequate HL levels, indicating a huge variation in how individuals were classified. This fact can cause serious biases in the measurement of individual and population HL levels, impacting the planning and effectiveness of interventions carried out by professionals and health services.

The results from table 3 demonstrated an interesting and new evidence related to the influence of the items on the formation of the score of HLS-EU-Q6. It was observed that the questions related to find information about how to deal with mental problems (Q3) and understand the information available in the media on how to stay healthier (Q6) had the greatest impact on the scores.

Some hypotheses can be considered to explain this fact, such as the increase in the incidence of mental problems in the world population, also mediated by aspects related to fake news and infodemics in social networks (WORLD HEALTH ORGANIZATION, 2019; BORGES DO NASCIMENTO IJ, et al., 2022).

It is pertinent to emphasize that it is difficult to compare the classification of the Brazilian sample with other studies due to the huge differences between the techniques adopted. We must bear in mind that the score of a scale it is not a continuous data or restricts or eliminates the possibility of using the mean as a reference for the analyses.

The set of techniques adopted in the present study enabled the combination of descriptive and distribution techniques with techniques whose property is to classify individuals into clusters (groupings). Thus, testing more than one possibility of initial classification and subsequent assessment of the precision of the cuts adopted protects against possible biases that may be created by the imprecision of the technique adopted and prevents arbitrary options. This combination to improve the accuracy also prevents or reduces the effects of asymmetries that can affect cutoffs.

It is important to remember that the cutoffs are made by percentiles and not by the score, and the score is obtained from the position of individuals in the distribution. An example is that we may have a score that runs through more than one quartile and, if this occurs, individuals with the same score on the instrument may be in different quartiles and would present classifications in different ranges. Discriminant analysis was used to prevent this type of inaccuracy from occurring.

Another fundamental aspect is to understand that the normality of the distribution is not the rule (PINO FA, 2014) and that asymmetry can decisively affect the analyzes and consequently the interpretation of the data (VERNA JP and ABDEL-SALAM ASG, 2019).

Despite the potential of the current normative data, this study present some limitations. The sample consisted of users of primary care who may have a different characteristic from users of private services. Data were collected only in one city. Future studies are needed for comparison of current normative data of HLS-EU-Q6 with other populations.
CONCLUSION

This study demonstrated that the HLS-EU-Q6 classification criteria to be used in Brazilian adults should be different from those proposed by the authors of the original study, so as not to generate biases in the measurement of health literacy levels in this population. The normative data of the Brazilian Portuguese version of the instrument indicated that the classification model with 3 specific ranges were able to correctly classify the participants according to their HL levels.

REFERENCES