

KnowBe4 Technical Documentation for the Security Awareness Proficiency Assessment (SAPA)

This document describes the process used by KnowBe4 to develop, test and validate the Security Awareness Proficiency Assessment (SAPA). To ensure the integrity of the assessment, KnowBe4 does not provide copies of the items¹.

What is SAPA and how does it work?

SAPA was introduced in 2019 as a method to measure an organization's proficiency across seven knowledge areas of security awareness. The assessment provides the organization with an overall security awareness benchmark and an ability to pinpoint weaknesses in the individual and the organization. To derive, individuals at the organization are surveyed and asked to answer 21 questions (three questions of varying difficulty per each of the seven knowledge areas). For each of the possible 21 questions that an end-user sees, there are at least two more like-items available in our question bank. This means the probability of anyone receiving the same set of questions again is more than **18,600 million to one**.

The order in which the questions and their answer choices are presented are randomized, further reducing the chances of two employees at the same organization receiving the exact same assessment. The chances of an employee getting the exact same assessment twice are infinitesimal. We cannot provide copies of the items nor approximate what items a particular user will receive.

We consider this to be a valid and valuable assessment for the purposes of determining a user's need for training. **The assessment can be used multiple times to re-assess a user's skills and security awareness proficiency over time.**

Technical Development

KnowBe4 prides itself on providing its customers with the freshest content of the highest quality. Based on a synthesis² of the latest research available, three years of customer feedback and extensive analyses of the data collected from over 3 million completions, we updated SAPA in 2022 with a few simple adjustments.

Steps 1-6 below describe the updates that were completed as part of a rigorous and continuous quality assurance review process by a team of subject matter experts, overseen by our assessment specialist.

¹ **Items** are questions or agreement statements. Although there is a larger bank of items available, SAPA presents only 21 items (questions) to the end-user with each iteration.

² **Research synthesis** is the process of combining, aggregating, integrating and synthesizing the results of multiple primary research studies aimed at testing the same conceptual hypothesis.

Step one

The revision process began with a detailed item-level analysis of all **63 items** in circulation³ within our existing dataset of over 3 million completions of the existing assessment. The difficulty level and discrimination index⁴ values of each item were recalculated, and the overall suitability of each item (against the new domain and subdomain criteria) was re-evaluated.

As a result, the difficulty level of 10 items was adjusted, 14 items were rewritten, eight items were removed, and **34 new items** (including two new fixed items) were added. The new fixed⁵ items provided a ranking based on the respondents' years of cybersecurity experience and their self-reported skill level, which was used as part of quality testing during the piloting stages (steps 3-6).

Step two

All new or updated items were first reviewed by two quality assurance (QA) team members and then critically evaluated for accuracy, clarity and suitability by a panel of subject matter experts. The panel included assessment specialists, instructional designers, linguists and multiple security experts. Any items that did not reach their high standards were either discarded or revised according to the panel's recommendations. All revisions were reviewed by the QA specialists again and re-evaluated by the panel. At the end of the final round, **76 items** remained.

Step three

We conducted an internal pilot (**n₁=116**)⁶ of the remaining 76 items and subjected the assessment to a series of rigorous validity tests. Controls were put in place to ensure that participants were randomly selected globally, from different areas of the business, and included employees with different lengths of employment (ranging from less than a week through to 5+ years) to ensure their exposure to the industry was varied. No employees involved in any stage of the assessment's development were included, and participation was anonymous. Using the pilot results, each item was analyzed and only items that scored well on the discrimination index (i.e., $DI \geq 0.4$) remained unchanged.

Step four

A second pilot was conducted externally on the revised set of **72 items** under similar conditions (**n₂=104**). All items performed well on the discrimination index (**0.493 \geq DI \geq 0.956**). We also calculated the Pearson product-moment correlation coefficient⁷ between the overall score of the pilot group and their self-assessment

³ These items have already undergone extensive quality testing and passed assurance. Previous external validity tests (performed when SAPA was first released, in 2019) revealed a strong/moderate positive correlation (**r₀ = 0.68**).

⁴ The **discrimination index** (DI) measures how discriminating items in an exam are. In other words, how well an item can differentiate between good candidates and less able ones. **DI values** range from -1 to +1, where values above 0.40 are considered by statisticians to be very discriminating (very good).

⁵ A **fixed item** is a question that is asked to every end-user. It is not randomly selected from a larger bank of items.

⁶ Also referred to as the sample size, n denotes the number of participants in the pilot. In the event there are multiple pilots, the pilots are numbered consecutively and the corresponding number is written in subscript (i.e., n₁, n₂, and n₃, refer to the first, second, and third pilots conducted).

⁷ In statistics, dependence or association is any statistical relationship, whether causal or not, between two variables. The **Pearson product-moment correlation coefficient** is a measure of the strength of a linear association between two variables (i.e., an individual's overall SAPA score and their cybersecurity knowledge) and is denoted by r.

of cybersecurity knowledge, which revealed a moderately strong correlation ($r_2=0.71$). With renewed confidence in individual item quality (based on the internal and external validity tests), we reviewed the difficulty level of the items next, paying particular attention to like-items⁸ to ensure there was an appropriate balance across the set. Any items that did not reach our high standards were further revised with minor edits or thrown out⁹. At the end of step four, **68 items** remained.

Step five

All 68 items were reviewed by two *different* quality assurance (QA) team members and the panel of Subject Matter Experts. A final review was completed by members of our data protection, legal and information security teams. **Two** items were flagged for minor edits, and **one** was removed.

Step six

In preparation for testing, the two fixed items (previously used for ranking purposes in the external validity tests and now no longer needed) were removed, leaving **65 items** in the question bank. Lastly, we assigned the assessment under real test conditions to the entire KnowBe4 workforce, excluding any employees that had been involved in the development of the assessment ($n_3 = 1498$).

A final quality check was done using the results of this test to verify individual item quality and ensure a balance in the difficulty of items across the different knowledge constructs being measured. After which, **57 items** remain and were localized for production.

Step seven

The edited Security Awareness Proficiency Assessment was rolled out to clients. Following a few months of usage, we analyzed the answers of $N>63,000$ users question by question. Five questions were identified as problematic based on item analysis of respondents. Subsequently, the analysts eliminated the five questions deciding that the question/answer set did not fit the expected statistical models for internal reliability, leaving 52 potential questions in the pool.

Further, the analysts determined that the rest of the SAPA was sufficiently sound as to move forward to production.

About the BETA version and SAPA localizations

This assessment **of 52 items** was published in February 2023 in US English after which localization commenced to update the assessment into all 34 of our standard supported languages¹⁰.

⁸ **Like-items**, sometimes also referred to as banking items, are items that reliably measure the same concept with the same level of validity as other items. They can be used interchangeably in an assessment or survey without affecting the quality of the overall result.

⁹ This is because like-items, by definition, must fulfill certain criteria to ensure they are as equally effective at measuring the same nuanced concept as each other (e.g., be rated at the same difficulty level) in order to be interchangeable with one another (i.e., "bankable").

¹⁰ For a list of the languages that KnowBe4 supports as standard, please see: <https://www.knowbe4.com/security-awareness-training-languages>

References:

- Edwards, Keith. "Examining the Security Awareness, Information Privacy, and the Security Behaviors of Home Computer Users." Ph.D diss., Nova Southeastern University, 2015. https://nsuworks.nova.edu/gscis_etd/947
- Hameed, M.A., Arachchilage, N.A.G. The role of self-efficacy on the adoption of information systems security innovations: a meta-analysis assessment. *Pers Ubiquit Comput* 25, 911–925 (2021). <https://doi.org/10.1007/s00779-021-01560-1>
- Mahabi, Victoria. "Information Security Awareness: System Administrators And End-User Perspectives At Florida State University." Ph.D diss., Florida State University, 2010. http://purl.flvc.org/fsu/fd/FSU_migr_etd-2798
- Parsons, Kathryn, Agata McCormac, Marcus Butavicius, Malcolm Pattinson, and Cate Jerram. "Determining Employee Awareness Using the Human Aspects of Information Security Questionnaire (HAIS-Q)." *Computers & Security* 42 (May 2014): 165-76. doi:10.1016/j.cose.2013.12.003.
- Parsons, Kathryn, Dragana Calic, Malcolm Pattinson, Marcus Butavicius, Agata McCormac, and Tara Zwaans. "The Human Aspects of Information Security Questionnaire (HAIS-Q): Two Further Validation Studies." *Computers & Security* 66 (May 2017): 40-51. doi:10.1016/j.cose.2017.01.004.

Veiga, Adéle Da. "An Approach to Information Security Culture Change Combining ADKAR and the ISCA Questionnaire to Aid Transition to the Desired Culture." *Information and Computer Security* 26, no. 5 (October 2018): 584-612. doi:10.1108/ics-08-2017-0056.