Title

Seeing the Light in Self-Reported Glare

Authors

Johnston, Elisha
Reise, Steven P
Spritzer, Karen L
et al.

Publication Date

2023

DOI

10.1027/1015-5759/a000798

Peer reviewed

Original Article

Seeing the Light in Self-Reported Glare

Elisha Johnston[1], Steven P. Reise[2], Karen L. Spritzer[3], and Ron D. Hays[3]

[1]Department of Molecular, Cellular, and Developmental Biology, University of California, Los Angeles, CA, USA
[2]Department of Psychology, University of California, Los Angeles, CA, USA
[3]Division of General Internal Medicine and Health Services Research (GIM-HSR), Department of Medicine, University of California, Los Angeles, CA, USA

Abstract: We conducted a secondary analysis to evaluate two glare items versus a composite score from the two items in a sample of 544 patients (38% women; *Mdn* = 29 years old; 13% high school education or less) before and after eye surgery. The first question was from a National Eye Institute (NEI) survey and the second question included a definition and picture of glare. At baseline, 28% of participants reported glare on the NEI item versus 39% on the question with the definition and picture. There was 76% agreement between the two questions (K = .46). Three months after baseline, there was no change in the percentage of participants who reported glare on the NEI question (27%), but a significant decrease in participants reporting glare on the question with the definition and picture (38% at baseline to 25% 3 months later). A 2-item glare scale was more reliable and highly correlated with multi-item measures of halos and starbursts than the individual items. Change in the scale was more strongly associated than the items with the change in satisfaction with vision and with satisfaction with surgery. The scale may provide a better indicator of glare than either item alone.

Keywords: glare, single-item measures, self-report, reliability

Multi-item scales are used in survey research to enhance the reliability of measurement (Hoeppner et al., 2011; Matthews et al., 2022). Internal consistency reliability (α) provides information about the associations among different items in a multi-item scale and increases with the number of items when correlations among the items are positive (Cronbach, 1951). That is, $\alpha = (k \times P_I)/(1 + (k - 1) \times P_I)$, where k is the number of items and P_I is the intraclass correlation. A reliability of .70 is a minimum threshold for group-level comparisons while .90 is used for individual-level assessment (Nunnally, 1978). Reliability is a necessary but not sufficient condition for valid measurement – the degree to which evidence supports the proposed use of a measure. Evidence for validity includes the extent to which associations of scores from the measure with other variables are consistent with hypotheses (American Educational Research Association et al., 2014).

Respondent burden and cost of administration increase with the number of items administered and the incremental value of additional items varies by the breadth of the concept being measured. Allen and colleagues (2022) noted that a single-item measure can be sufficient, particularly when the construct is unambiguous or narrow in scope. Robins and colleagues (2001) showed that a single-item measure of self-esteem performed like the 10-item Rosenberg Self-Esteem Scale. Similarly, West and colleagues (2012) demonstrated that single items perform similarly to the Maslach Burnout Inventory long-form measures of emotional exhaustion and depersonalization in terms of associations with suicidality, serious thoughts of dropping out of medical school, endorsing dishonest behavior, disagreeing with an altruistic attitude, and perceived major medical error.

Measuring ocular symptoms such as glare is important in evaluating conditions such as near-sightedness, cataracts, or macular degeneration (Cui et al., 2021). The brightness acuity test (BAT) has been used to evaluate disability glare in clinical practice, but the BAT does not reflect day-to-day experience with glare (Aslam et al., 2007). Researchers have often used single self-report items to assess glare (Chiam et al., 2006; Häring et al., 2001; Paz et al., 2013; Vingolo et al., 2007).

Multi-item scales have also been used to assess glare. The 4-item daily vision glare scale asks about difficulty driving at night with oncoming headlights, seeing faces across the street in bright sunlight, reading numbers on the television, and playing cards using five response options (*do not do the activity because of vision problems*; *extremely difficult*; *moderately difficult*; *a little difficult*; *not difficult at all*). The internal consistency reliability of this daily vision glare scale was only .63 (Valbuena et al., 1999). The 2-item National Eye Institute (NEI) Refractive Error Quality of Life glare scale asks if one has experienced glare in the last 7 days and how often when one is around bright lights at

night you see starbursts or halos that bother you or make it difficult to see. Test-retest (median of 16 days between assessments) reliability for this 2-item scale was only .64 (Hays et al., 2003). In summary, prior work indicates noteworthy challenges in the measurement of glare.

Because of the limitations of existing measures of glare, investigators on the Patient-Reported Outcomes with Laser in situ keratomileusis surgery study (PROWL) developed a new item: "In the last 7 days, have you noticed any glare?". To standardize the meaning of glare to respondents, the new question included a definition and a picture depicting glare (see Figure 1). For comparison to previous work, the PROWL investigators also included the NEI glare item: "Have you experienced glare in the last 7 days?" (Hays et al., 2003). Given the similarity in wording of the NEI and the new PROWL item, one might expect to find consistent answers to these questions.

We compare responses to these two self-report items to examine if using one item to assess glare from the patient's perspective provides similar information to that of the composite score created from both items. In doing so, we address three research questions:

Research Question 1 (RQ1): What is the association between two similarly worded self-report items assessing the same construct?

Research Question 2 (RQ2): How much more reliable is the 2-item glare scale than the individual items?

Research Question 3 (RQ3): Does the 2-item scale correlate more strongly than the single-item with multi-item indicators of glare, satisfaction with vision, ocular surface disease, and satisfaction with surgery?

Methods

Design

This is a secondary analysis of two similar items administered to patients with myopia, hyperopia, or astigmatism (Hays et al., 2017). Patients planning to undergo LASIK surgery were screened and enrolled by investigators at several sites: San Diego and Stanford, California; Indianapolis, Indiana; Overland Park, Kansas; Baltimore, Maryland; and Sioux Falls, South Dakota. Sites performed the LASIK procedure and postoperative care based on the clinical judgment of the surgeons. The PROWL study is registered at Clinicaltrials.gov (NCT0152629 and NCT01655420).

A total of 564 adults were recruited for the PROWL study: 252 active-duty US Navy service personnel and 312 adult civilians. Survey data were available for 544 individuals at baseline (96% of those enrolled in the study): 68% White, 10% Asian, 6% Black, 11% Hispanic, and 5% other; 38% women; *Mdn* = 29 years old (range 21–57); 13% high school education or less. Data were available at baseline (before surgery) and 1 month later for 495 patients (88% of those enrolled) and 3 months later for 480 patients (85% of those enrolled). Data collection began in August 2011 and was completed on June 27, 2014.

Questionnaires were self-administered via the Web. Data collection with civilians was conducted under the US Food and Drug Administration Research Involving Human Subjects Committee, a central Institutional Review Board (IRB) for some sites, and university IRBs for others. The protocol for data collection with the military personnel was approved by the Naval Medical Center San Diego IRB in compliance with all applicable federal regulations governing the protection of human participants. All participants provided written as well as oral informed consent.

Glare Items

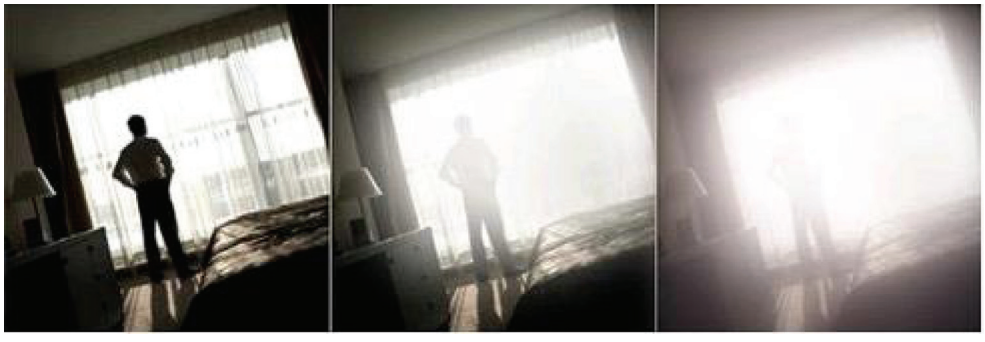
As noted above, the study included the NEI item: "Have you experienced glare in the last 7 days?" (*Yes/No*). It was the 33rd question on the 161-item baseline (preoperative) questionnaire. In addition, a second question was included (see Figure 1): "In the last 7 days, have you noticed any glare?" (*Yes, but only when not wearing glasses or contact lenses/Yes, but only when wearing glasses or contact lenses/Yes, when wearing and when not wearing glasses or contact lenses/No, not at all*). This second question was the 50th question on the baseline questionnaire.

Self-Report Measures Included to Assess Construct Validity

Halos and starbursts are examples of glare, and the study administered an 8-item halos scale and an 8-item starbursts scale (see: https://www.fda.gov/media/101491/download). The 8-item scales are scored on a 0–100 possible range with a higher score being better (fewer symptoms). Internal consistency reliability was .97 for both scales (Hays et al., 2017). The study also included the 8-item Ocular Surface Disease Index (Schiffman, 2000) that assesses whether one experiences eyes that are sensitive to light, eyes that feel gritty, painful, or sore, blurred vision, and poor vision. Internal consistency reliability for the index was .91 in a clinical trial of 571 adults with blepharitis (Hosseini et al., 2018). Also included was an item assessing satisfaction with vision and an 8-item satisfaction with surgery scale, both scored on a 0–100 possible range with higher scores indicating more satisfaction. The satisfaction with vision item had a test-retest (*Mdn* = 11 days) reliability of .67 while the satisfaction with surgery scale had an internal consistency reliability of .90 (Hays et al., 2017).

INSTRUCTIONS: The next few questions are about *glare*. By glare, we mean *difficulty seeing well when there are bright lights* like headlights or sunlight, such as shown in the images below. These images may not represent exactly what you see and your symptoms may be more or less severe than what is shown.

No glare ⟶ Severe glare



In the last 7 days, have you noticed any glare?

☐ Yes, but ONLY when NOT wearing glasses or contact lenses

☐ Yes, but ONLY when wearing glasses or contact lenses

☐ Yes, when wearing AND when not wearing glasses or contact lenses

☐ No, not at all

Figure 1. Glare question that includes a picture.

We hypothesized that reports of glare would be positively associated with the halos and starbursts scales at baseline and that increases in glare from baseline to the 1-month and 3-month follow-ups would be negatively correlated with satisfaction with vision and the satisfaction with surgery scale.

Analysis Plan

Research Question 1: Association Between the Two Glare Questions

Because the NEI item is dichotomous, we collapsed the first three options for the second glare question (i.e., "picture" item) into a single yes category to make it comparable to the NEI glare item. We report the percentage of patients reporting glare on each item, the percentage agreement, product-moment correlations between the items, and the kappa coefficient of agreement (Cohen, 1960) between the items at the baseline of the study. In addition, we compare changes in glare for each item from baseline to 1 month and 3 months later.

Research Question 2: Estimated Reliability of Glare Questions and 2-Item Glare Scale

We examine mean scores on the halos and starburst scales for all possible combinations of responses to the two glare items: (1) No on both; (2) Yes on question 1 only, (3) Yes on question 2 only, and (4) Yes on both questions. Then, we estimate the internal consistency reliability of the 2-item simple summated glare scale. To estimate the reliability of the single items, we apply the Spearman-Brown (Clark, 1935) prophecy formula: $n \times \alpha/(1 + (n - 1)\alpha)$, where n is $1/k$, and k is the number of items in the scale (i.e., $n = 1/2$).

We also estimate test-retest reliability over a mean of 11 days for a random subsample of 50 individuals with product-moment correlations. To reflect any mean change, we also estimate intraclass correlations using the two-way random effects model: $N\ (MS_{between} - MS_{within})/(N\ MS_{between} + MS_{time} - MS_{within})$.

Research Question 3: Associations of Glare Items and Scale With Other Variables

We estimate Spearman rank-order correlations of the 2-item glare scale with the halos and starburst scales and with the Ocular Surface Disease Index. Then, we report product-moment correlations of change in the two glare items and change in the 2-item glare scale with change in satisfaction with vision and satisfaction with surgery at the two follow-ups. We evaluate differences in correlations using Steiger's (1980) *z*-statistic.

Results

Research Question 1: Association Between the Two Glare Questions

Study Baseline

As shown in Table 1, fewer of the 544 patients reported glare in the last 7 days on the NEI item (28%) than on the picture glare item (39%) at the baseline of the study. There was 76% agreement between the two dichotomously scored questions, r = .48, and K = .46, indicating "fair" (Fleiss, 1981) or "moderate" (Landis & Koch, 1977) agreement. This level of agreement was less than one might expect for two similarly worded items.

Table 1. Cross-tabulation of two questions about glare

	(1) Have you experienced glare in the last 7 days?		
(2) In the last 7 days have you noticed any glare?	No	Yes	Total
No	**297**	35	332
Yes	97	**115**	212 (39%)
Total	394	150 (28%)	544

Note. Question 1 is NEI-RQL item and question 2 is an item that includes a picture. Bold entries represent agreement between responses to the 2 items.

One Month and Three Months Change in Self-Reported Glare

Figure 2 shows patterns of self-reported glare over time for those with data at baseline (before surgery), 1 month later (n = 495), and 3 months later (n = 480). Note that glare at baseline was one percentage point lower for the subset with longitudinal data than for the slightly larger baseline sample noted above. Among those with data at both time points, there was a significant increase in reported glare

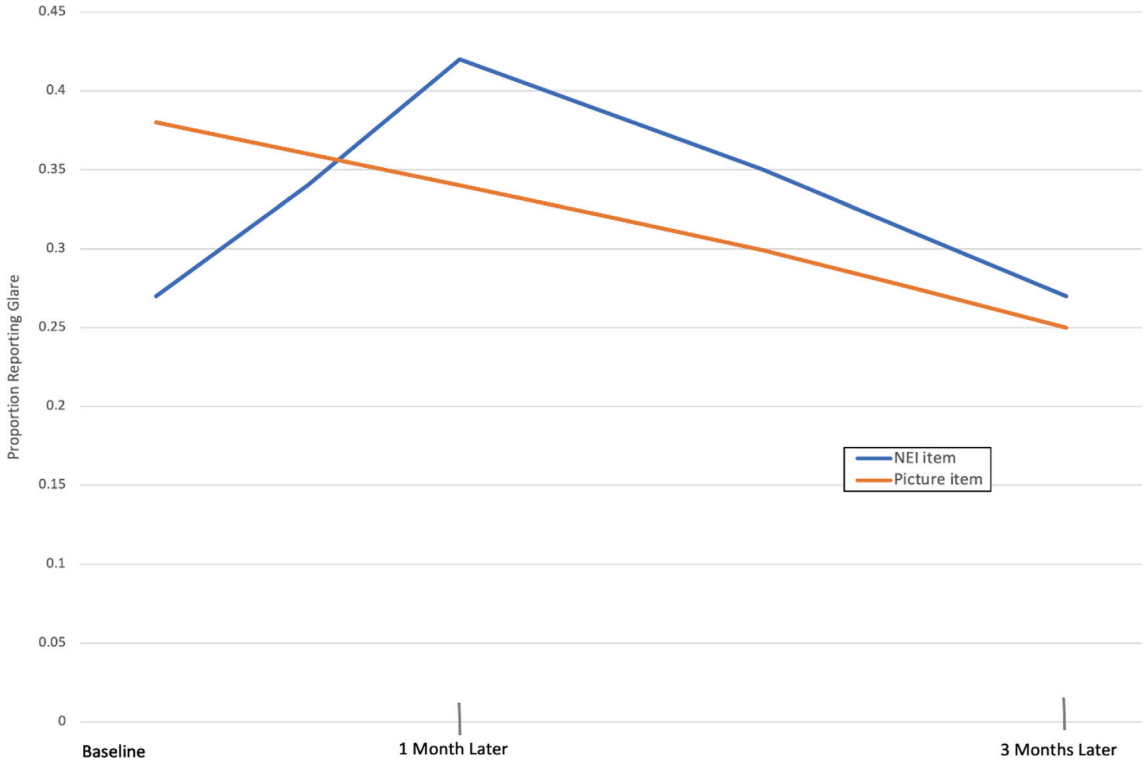


Figure 2. Reported glare at the three-time points for two survey items.

Table 2. Means scores on satisfaction with current vision item, halos, and starbursts by combination of responses to two glare items

	No for both glare questions	Yes for Question 1 only	Yes for Question 2 only	Yes for both glare questions
Halos	90[a]	69[b]	63[b,c]	56[c]
Starbursts	85[a]	57[b]	60[b]	53[b]

Note. Question 1: Have you experienced glare in the last 7 days? Question 2: In the last 7 days have you noticed any glare? The 8-item halos and 8-item starbursts scales are scored on a 0–100 possible range with a higher score being better (fewer symptoms). ANOVA F-statistic = 74.73 ($p < .0001$) for halos and 60.47 ($p < .0001$) for starbursts. Means in a row that share a superscript letter do not differ significantly from one another on Duncan's multiple range test.

Table 3. Product–moment correlations (p-values) of change in glare items and glare scale with change in satisfaction with vision and satisfaction with surgery

	Month 1 – Baseline	Month 1	Month 3 – Baseline	Month 3
	Satisfaction with vision	Satisfaction with surgery	Satisfaction with vision	Satisfaction with surgery
Change in NEI item	−.14 (.0025)	−.20 (<.0001)	−.25 (<.0001)	−.31 (<.0001)
Change in picture item	−.18 (<.0001)	−.22 (<.0001)	−.18 (<.0001)	−.23 (<.0001)
Change in 2-item glare scale	−.19 (<.0001)	−.25 (<.0001)	−.25 (<.0001)	−.31 (<.0001)

Note. Listwise deletion of cases $n = 460$. Satisfaction with vision is a single item. Satisfaction with surgery is an 8-item scale. Both are scored on a 0–100 possible range, with 100 representing more satisfaction.

from 27% at baseline to 42% at the 1-month follow-up on the NEI question ($t(494) = 5.50$, $p < .0001$) but reported glare on the picture question did not change significantly (38–34%; $t(494) = -1.72$, $p = .0874$). Three months later compared to baseline, there was no significant change for the NEI question (27% at both time points, $t(479) = 0.25$, $p = .8049$), but there was a significant decrease in reported glare on the picture question from 38% at baseline to 25% 3 months later ($t(479) = -5.01$, $p < .0001$). These results suggest inconsistent conclusions about the change in glare over time.

The percentage of those reporting glare on either or both items was: 45% at baseline, 51% at 1 month post-baseline, and 33% at 3 months post-baseline. The increase in glare reported on either item from baseline to 1 month later was almost significant ($t(491) = 1.94$, $p = .0531$) and the decrease from baseline to 3 months later was significant ($t(460) = -3.96$, $p < .0001$).

Research Question 2: Estimated Reliability of Glare Questions and 2-Item Glare Scale

Table 2 presents mean 0–100 scores for the halos and starbursts scales by the four cells that fully represent possible responses to the two glare questions: (1) "no" on both ($n = 297$), (2) "yes" on question 1 only ($n = 35$), (3) "yes" on question 2 only ($n = 97$), and (4) "yes" on both questions ($n = 115$). Halos and starbursts mean scale scores were not significantly different for answers of yes to only one of the glare questions, providing support for simple-summated scoring of the 2-item glare scale: 0 = *no glare reported for both items*, 1 = *glare reported on one of the items*, and 2 = *glare reported on both items*.

Internal consistency reliability was .64 at both baseline and 1 month later. Using the Spearman-Brown formula, the estimated reliability for a single item, $P_I = 0.48$. Product-moment (intraclass) test-retest correlations were $r = .55$ ($P_I = 0.52$) for the NEI item, $r = .54$ ($P_I = 0.54$) for the pictured item, and $r = .61$ ($P_I = 0.61$) for the 2-item glare scale. These reliability estimates, all below the .70 minimum reliability threshold for group-level comparisons (Nunnally, 1978), provide further evidence of unique information in responses to the two glare items.

Research Question 3: Associations of Glare Items and Scale With Other Variables

Spearman correlations of the NEI item, picture item, and the 2-item scale, respectively were $r_s = -.39$, $-.50$, and $-.53$ with the 8-item halos scale, $r_s = -.39$, $-.43$, and $-.49$ with the 8-item starbursts scale, and $r_s = .34$, .37, and .42 with the Ocular Surface Disease Index. Product-moment correlations were similar and showed consistent differences in the magnitude of the associations.

Product-moment correlations of the two single items and the 2-item glare scale with change in satisfaction with vision and the 8-item satisfaction with vision scale are provided in Table 3. Change from baseline to the 1-month follow-up in the 2-item glare scale was significantly more strongly associated than change in the NEI item with change in satisfaction with vision ($z(457) = 2.03$, $p = .042$) and satisfaction with surgery at 1 month ($z(457) = 2.16$, $p = .030$). Change from baseline to the 3-month follow-up in the 2-item glare scale was significantly more strongly associated than was the change in the picture item with change in satisfaction with

vision and satisfaction with surgery at 3 months ($z(457) = 2.81$, $p = .005$ and $z(457) = 3.56$, $p < .001$, respectively).

Discussion

The NEI and the picture glare items correlate as highly with one another ($P_I = .48$) as items in many multi-item scales. The 2-item glare scale had an internal consistency reliability of .64. A 3-item scale with the same intraclass correlation would have an internal consistency reliability of .73.

The prevalence of glare reported at baseline on the NEI item was much lower than glare reported on the picture item, but the proportion of study participants reporting glare 1 month later on the NEI item increased. In contrast, the amount of glare 1 month later for the picture item was not significantly different than the level reported at the baseline of the study. The proportion of people with these symptoms also did not change from baseline to 1 month later (Hays et al., 2017). The use of definitions and pictures could account for the similar lack of significant change for these measures from baseline to 1 month later. Change from baseline to 3 months later also differed for the two items. The estimate of glare for the NEI item was not significantly different from the baseline, but the glare was reported to be significantly less common on the picture item 3 months after surgery compared to the baseline.

Which way of asking about glare should be used in the future? Further investigation (e.g., cognitive interviews) could shed light on differences in reports for these two glare items. However, responses to the two items in this study suggest that reporting glare on both items is more indicative of glare than reporting glare on only one of the items. Hence, the 2-item glare scale represents three levels of increasing glare:

- No glare reported on either item.
- Glare reported on one of the items.
- Glare reported on both items.

Use of this 2-item scale is a better option than either glare item alone because it is more reliable and more strongly associated with multi-item scales assessing similar constructs (halos, starbursts, ocular surface disease) at baseline, and with change in satisfaction with vision and change in satisfaction with surgery at the 1-month follow-up. The trend toward an increase in glare 1 month after and a decrease 3 months after LASIK based on reporting glare on either item is consistent with a study of 185 patients before and four months after surgical correction of myopic refractive error that used the same NEI item (McDonnell et al., 2003).

How does this study contribute to the literature on the use of single versus multiple items to assess a construct? The fact that single items can adequately represent some concepts for some applications (Allen et al., 2022; Robins et al., 2001; West et al., 2012) is the underpinning of short-form measures (Hays et al., 1991; Ware & Sherbourne, 1992). Within the limits of the reliability of the full item set, a subset of items can be selected that yields whatever target level of reliability is desired for a particular application. For scales that represent narrow concepts, the items will tend to be more highly correlated, and the true score may be estimated accurately with fewer items. In addition, it is possible for a multi-item scale to have a single item that is very highly associated with the scale total and to perform like the full-length scale. The lower response burden makes parsimonious measures desirable when they retain the psychometric strengths of longer measures. However, the equivalence of scores produced from single versus multiple items needs to be demonstrated and tradeoffs carefully considered in different circumstances.

Limitations and Strengths

This secondary analysis has some limitations. The data were limited to surveys administered in English. In addition, it is unclear how well the results generalize to other samples and contexts. There was also some attrition of study participants over time, as is the case in any longitudinal study. However, the study had multiple strengths. For example, the use of pictures to depict glare for the second item represents the state of the science for assessing visual symptoms. In addition, collecting data over three time points provided information about changes in reported symptoms associated with the LASIK intervention.

Conclusions

With respect to the more general question of how much is lost when using single rather than multiple items, this study suggests some added value in using two self-report items rather than one item in assessing a complex symptom. The NEI item relies on the respondent's own definition of glare. In contrast, the second item uses a definition and picture to help standardize what is meant by glare. The items yielded conflicting information about changes in glare associated with LASIK surgery. The NEI item indicated an increase in glare 1 month after surgery, but no change 3 months later compared to baseline. The picture item indicated a significant decrease in glare 3 months after surgery. The use of both items can lead to greater reliability and higher accuracy of estimation of glare than either item alone. The percentage of those reporting glare on either or both items was: 45% at baseline, 51% at 1-month post-baseline, and 33% at 3-month post-baseline.

While the results indicate potential benefits for using the NEI and picture glare items together, further investigations are needed to provide additional information about the two items. Results may not generalize to other contexts. Results may differ for other measures. The clearer and narrower the concept, the more likely it is that a single item will be sufficient to assess it.

References

Allen, M. S., Iliescu, D., & Greiff, S. (2022). Single item measures in psychological science. *European Journal of Psychological Assessment, 38*(1), 1–5. https://doi.org/10.1027/1015-5759/a000699

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Aslam, T. M., Haider, D., & Murray, I. J. (2007). Principles of disability glare measurement: An ophthalmological perspective. *Acta Ophthalmologica Scandinavica, 85*(4), 354–360. https://doi.org/10.1111/j.1600-0420.2006.00860.x

Chiam, P. J. T., Chan, J. H., Aggarwal, R. K., & Kasaby, S. (2006). Restor intraocular lens implantation in cataract surgery: Quality of vision. *Journal of Cataract and Refractive Surgery, 32*(9), 1459–1463. https://doi.org/10.1016/j.jcrs.2006.04.015

Clark, E. L. (1935). Spearman-brown formula applied to ratings of personality traits. *Journal of Educational Psychology, 26*(7), 552–555. https://doi.org/10.1037/h0054722

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555

Cui, Q. N., Hays, R. D., Tarver, M. E., Spaeth, G. L., Paz, S. H., Weidmer, B., Fellman, R. L., Vold, S. D., Eydelman, M., & Singh, K. (2021). Vision-targeted health-related quality-of-life survey for evaluating minimally invasive glaucoma surgery. *American Journal of Ophthalmology, 229*, 145–151. https://doi.org/10.1016/j.ajo.2021.03.064

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). The measurement of interrater agreement. In W. A. Shewart & S. S. Wilks (Eds.), *Statistical methods for rates and proportions* (3rd ed., pp. 598–626). John Wiley & Sons, Inc. https://doi.org/10.1002/0471445428

Häring, G., Dick, B. H., Krummenauer, F., Weissmantel, U., & Kröncke, W. (2001). Subjective photic phenomena with refractive multifocal and monofocal intraocular lenses. *Journal of Cataract and Refractive Surgery, 27*(2), 245–249. https://doi.org/10.1016/s0886-3350(00)00540-x

Hays, R. D., Larson, C., Nelson, E. C., & Batalden, P. B. (1991). Hospital quality trends: A short-form patient-based measure. *Medical Care, 29*(7), 661–668. https://doi.org/10.1097/00005650-199107000-00006

Hays, R. D., Mangione, C. M., Ellwein, L., Lindblad, A. S., Spritzer, K. L., & McDonnell, P. J. (2003). Psychometric properties of the National Eye Institute – Refractive Error Quality of Life instrument. *Ophthalmology, 110*(12), 2292–2301. https://doi.org/10.1016/j.ophtha.2002.07.001

Hays, R. D., Tarver, M. E., Spritzer, K. L., Reise, S., Hilmantel, G., Hofmeister, E. M., Hammel, K., May, J., Ferris, F., & Eydelman, M. (2017). Assessment of the psychometric properties of a questionnaire assessing Patient-Reported Outcomes with Laser in situ Keratomileusis (PROWL). *JAMA Ophthalmology, 135*(1), 3–12. https://doi.org/10.1001/jamaophthalmol.2016.4597

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305–312. https://doi.org/10.1016/j.jsat.2011.04.005

Hosseini, K., Bourque, L. B., & Hays, R. D. (2018). Development and evaluation of a measure of patient-reported symptoms of blepharitis. *Health and Quality of Life Outcomes, 16*(1), Article 11. https://doi.org/10.1186/s12955-018-0839-5

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Matthews, R. A., Pineault, L., & Hong, Y.-H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology, 37*(4), 639–673. https://doi.org/10.1007/s10869-022-09813-3

McDonnell, P. J., Mangione, C., Lee, P., Lindblad, A. S., Spritzer, K. L., Berry, S., & Hays, R. D. (2003). Responsiveness of the National Eye Institute refractive error quality of life instrument to surgical correction of refractive error. *Ophthalmology, 110*(12), 2302–2309. https://doi.org/10.1016/j.ophtha.2003.02.004

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Paz, S. H., Slotkin, J., McKean-Cowdin, R., Lee, P., Owsley, C., Vitale, S., Varma, R., Gershon, R., & Hays, R. D. (2013). Development of a vision-targeted health-related quality of life item measure. *Quality of Life Research, 22*(9), 2477–2487. https://doi.org/10.1007/s11136-013-0365-1

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*(2), 151–161. https://doi.org/10.1177/0146167201272002

Schiffman, R. M. (2000). Reliability and validity of the Ocular Surface Disease Index. *Archives of Ophthalmology, 118*(5), 615–621. https://doi.org/10.1001/archopht.118.5.615

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245–251. https://doi.org/10.1037/0033-2909.87.2.245

Valbuena, M., Bandeen-Roche, K., Rubin, G. S., Munoz, B., & West, S. K. (1999). Self-reported assessment of visual function in a population-based study: The SEE project. Salisbury Eye Evaluation. *Investigative Ophthalmology and Visual Science, 40*(2), 280–288.

Vingolo, E. M., Grenga, P., Iacobelli, L., & Grenga, R. (2007). Visual acuity and contrast sensitivity: AcrySof restor apodized diffractive versus ACRYSOF Sa60 at monofocal intraocular lenses. *Journal of Cataract and Refractive Surgery, 33*(7), 1244–1247. https://doi.org/10.1016/j.jcrs.2007.03.052

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*, 473–483.

West, C. P., Dyrbye, L. N., Satele, D. V., Sloan, J. A., & Shanafelt, T. D. (2012). Concurrent validity of single-item measures of emotional exhaustion and depersonalization in burnout assessment. *Journal of General Internal Medicine, 27*(11), 1445–1452. https://doi.org/10.1007/s11606-012-2015-7

History
Received January 13, 2023
Revision received August 7, 2023
Accepted August 12, 2023
Published online November 10, 2023
EJPA Section / Category Methodological topics in assessment

Conflict of Interest
The authors report no conflicts of interest in association with this manuscript.

Publication Ethics
All procedures in studies involving human participants were performed in accordance with the ethical standards of the US Food and Drug Administration Research Involving Human Subjects Committee, a central Institutional Review Board (IRB) for some sites, and university IRBs for others. The protocol for data collection with the military personnel was approved by the Naval Medical Center San Diego IRB.

Authorship
Elisha Johnson – Writing, original draft, review and editing, formal analysis; Steven P. Reise – Writing, original draft, review and editing, data curation, formal analysis; Karen L. Spritzer – Writing, original draft, review and editing, data curation, formal analysis; Ron D. Hays – Conceptualization, writing, original draft, review and editing, data curation, formal analysis. All authors approved the final version of the article.

Open Science
The US Food and Drug Administration, the National Eye Institute, and the Department of Defense launched the LASIK Quality of Life Collaboration Project that collected the data. The data are not publicly available.
Open Data: The information needed to reproduce all of the reported results is not openly accessible.
Open Materials: The information needed to reproduce all the reported methodologies is not openly accessible.
Preregistration of Studies and Analysis Plans: This study was preregistered at Clinicaltrials.gov (NCT0152629 and NCT01655420).

Funding
The Department of Health and Human Services (DHHS) and the Department of Defense provided funding for the data analyzed in this study. The DHHS and the Department of Defense funded and participated in the design and conduct of the study on this secondary analysis is based upon. The senior author was supported in part by the Resource Centers for Minority Aging Research Center for Health Improvement of Minority Elderly (RCMAR/CHIME) under NIH/NIA Grant P30-AG021684

ORCID
Ron D. Hays
https://orcid.org/0000-0001-6697-907X
Karen L. Spritzer
https://orcid.org/0000-0002-5969-0216
Steven P. Reise
https://orcid.org/0000-0002-5408-6992

Ron D. Hays
Division of General Internal Medicine and
Health Services Research (GIM-HSR)
Department of Medicine
University of California, Los Angeles
1100 Glendon Avenue, Suite 850
Los Angeles, CA 90024
USA
drhays@ucla.edu

