

King-Devick Test Reliability in National Collegiate Athletic Association Athletes: A National Collegiate Athletic Association–Department of Defense Concussion Assessment, Research and Education Report

Katherine M. Breedlove, PhD, ATC*; Justus D. Ortega, PhD†; Thomas W. Kaminski, PhD, ATC, FNATA, FACSM‡; Kimberly G. Harmon, MD§; Julianne D. Schmidt, PhD, ATC||; Anthony P. Kontos, PhD¶; James R. Clugston, MD, MS#; Sara P. D. Chrisman, MD, MPH**; Michael A. McCrea, PhD††; Thomas W. McAllister, MD‡‡; Steven P. Broglio, PhD, ATC, FNATA§§; Thomas A. Buckley, EdD, ATC‡

*Center for Clinical Spectroscopy, Brigham and Women's Hospital and Department of Radiology, Harvard Medical School, Boston, MA; †Department of Kinesiology and Recreation Administration, Humboldt State University, Arcata, CA; ‡Department of Kinesiology and Applied Physiology, University of Delaware, Newark; §Department of Family Medicine, University of Washington, Seattle; ||Department of Kinesiology, University of Georgia, Athens; ¶Department of Orthopaedic Surgery, University of Pittsburgh, PA; #Department of Community Health and Family Medicine, Department of Neurology, and Division of Sports Health, University Athletic Association, University of Florida, Gainesville; **Seattle Children's Research Institute and Department of Pediatrics, University of Washington, Seattle; ††Department of Neurosurgery, Medical College of Wisconsin, Milwaukee; ‡‡Department of Psychiatry, Indiana University, Indianapolis; §§School of Kinesiology, University of Michigan, Ann Arbor

Context: The King-Devick (KD) test has received considerable attention in the literature as an emerging concussion assessment. However, important test psychometric properties remain to be addressed in large-scale independent studies.

Objective: To assess (1) test-retest reliability between trials, (2) test-retest reliability between years 1 and 2, and (3) reliability of the 2 administration modes.

Design: Cross-sectional study.

Setting: Collegiate athletic training facilities.

Patients or Other Participants: A total of 3248 intercollegiate student-athletes participated in year 1 (male = 55.3%, age = 20.2 ± 2.3 years, height = 1.78 ± 0.11 m, weight = 80.7 ± 21.0 kg) and 833 participated in both years.

Main Outcome Measure(s): Time, in seconds, to complete the KD error free. The KD test reliability was assessed between trials and between annual tests over 2 years and stratified by test modality (spiral-bound cards [n = 566] and tablet [n = 264]).

Results: The KD test was reliable between trials (trial 1 = 43.2 ± 8.3 seconds, trial 2 = 40.8 ± 7.8 seconds; intraclass

correlation coefficient [ICC] (2,1) = 0.888, $P < .001$), between years (year 1 = 40.8 ± 7.4 seconds, year 2 = 38.7 ± 7.7 seconds; ICC [2,1] = 0.827, $P < .001$), and for both spiral-bound cards (ICC [2,1] = 0.834, $P < .001$) and tablets (ICC [2,1] = 0.827, $P < .001$). The mean change between trials for a single test was -2.4 ± 3.8 seconds. Although most athletes improved from year 1 to year 2, 27.1% (226 of 883) of participants demonstrated worse (slower) KD times (3.2 ± 3.9 seconds) in year 2.

Conclusions: The KD test was reliable between trials and years and when stratified by modality. A small improvement of 2 seconds was identified with annual retesting, likely due to a practice effect; however, 27% of athletes displayed slowed performance from year 1 to year 2. These results suggest that the KD assessment was a reliable test with modest learning effects over time and that the assessment modality did not adversely affect baseline reliability.

Key Words: mild traumatic brain injury, vision, oculomotor, baseline testing, psychometric properties

Key Points

- The King-Devick test was a reliable test with a modest learning effect over time.
- The assessment modality did not seem to adversely affect baseline reliability.
- Improvements in time between trial 1 and trial 2 reinforce the recommendation that 2 trials are needed for a baseline score.
- Among otherwise healthy collegiate student-athletes, 27% scored worse (ie, “failed”) their year 2 baseline test, emphasizing the need for a multifaceted concussion-assessment battery.

The identification of concussion in athletes has become a public health concern, with more than 4 million US emergency department visits annually.¹ Early detection of concussion improves patient outcomes; thus, a valid, sensitive, and easily administered assessment battery is a vital component of the sideline examination.^{2,3} The most popular sideline examination is the evolving version of the Sport Concussion Assessment Tool (SCAT5), which is a multifaceted assessment that tests cognition and balance and includes self-reported symptoms.⁴⁻⁶ However, visual disturbances, such as deficits in saccadic movement, accommodation, and convergence, are commonplace in both adolescents and adults and, therefore, visual assessments are typically abnormal postconcussion.⁷⁻⁹ The SCAT5 does not test the vestibular-ocular system; hence, a valid, reliable, and clinically feasible visual assessment may improve concussion management.⁴

The King-Devick (KD) test was created in 1976 by Alan King and Steven Devick¹⁰ to evaluate horizontal-saccade performance in relation to reading difficulties in children and was adapted for use with concussion screening¹¹ in 2011. A substantial number of the brain's circuits are involved in vision; as a result, an efficient visual evaluation is an appropriate addition to the multifaceted concussion-assessment battery.¹² The KD, which is a rapid number-naming assessment, is sensitive to vestibulo-ocular changes caused by concussion (sensitivity = 86%, specificity = 90%); the test involves complex cognitive function including visual-motor coordination, language function, and attention and can be completed in about 2 minutes.^{13,14} The authors¹⁴ of a systematic review identified a mean decrease (slowing) of 4.8 seconds on the KD test after a sport-related concussion compared with baseline, whereas nonconcussed athletes demonstrated a 1.9-second improvement in test performance. When the KD was added to the multifaceted concussion-assessment protocol of cognitive and balance testing and self-reported symptoms, the sensitivity improved to 100%, thereby supporting the inclusion of visual screening.^{15,16}

The most common concussion-management practice pattern for athletic trainers is to perform a baseline assessment when the student-athlete begins intercollegiate athletics and then reassess after a suspected concussion by comparing the results.⁴⁻⁶ Thus, the reliability of a test over time is a critical component of its efficacy, as common cognitive and balance tests demonstrate improvements with repeated administrations over time, likely secondary to a practice effect.^{17,18} Similarly, the KD has an inherent and expected learning effect with mean improvements of 2.8 seconds (faster) between 2 baseline trials and up to 6.4 seconds improvement between test administrations during the season.^{14,19,20} The baseline KD test consists of 2 trials. Because of the repetitive nature of the test, clinicians have anecdotally questioned the need for the second trial of the baseline test. No prior researchers have investigated the between-trials reliability of the KD test to ascertain if 2 trials are necessary for a baseline score.

Overall, the reliability of the KD test in healthy individuals has ranged from moderate to excellent (intra-class correlation coefficient [ICC] = 0.74–0.92) in boxers, mixed martial arts fighters, and collegiate student-athletes; however, many of these tests were administered over relatively short periods of time.^{11,14,20} Less is known about

the year-to-year test-retest reliability, which is important because the test score after a suspected concussion may be compared with a baseline score obtained years prior.^{5,6} In a recent large study²⁰ of National Collegiate Athletic Association (NCAA) student-athletes, reliability was only moderate (ICC = 0.74), with a 2.9-second improvement in the second year of testing, which raises concerns about test efficacy. Furthermore, current findings^{21,22} suggested that both reading level and test modality influenced performance on the KD test. Finally, the KD test has recently undergone a methodologic change from spiral-bound cards to a computerized tablet version. Early results²³ indicated differences in performance time between the modalities, which might also translate to differences in reliability.

The KD test has received considerable attention in the literature as part of an emerging concussion-assessment approach.²⁴ However, important test psychometric properties remain to be addressed in large-scale independent studies. Despite the high reliability, the manufacturer recommends that annual baseline assessments be performed in individuals age 10 and older, which may become time consuming and expensive when entire teams or athletic departments must be evaluated.²⁵ Although preliminary annual reliability data from the NCAA-Department of Defense (DoD) Concussion Assessment, Research and Education (CARE) data set were previously presented,²⁶ our aim was to expand on these findings by including important clinical considerations. Therefore, we had 3 goals: to assess (1) test-retest reliability between trials, (2) test-retest reliability between years 1 and 2, and (3) reliability stratified by administration modality (spiral-bound cards or tablet). We hypothesized that the KD would have good reliability (ICC > 0.75), but a clear practice effect would be present with repeat administration.

METHODS

Participants

Participants (N = 3248) were recruited from an ongoing prospective cohort study, the CARE Consortium, sponsored by the NCAA and the DoD (Tables 1 and 2). Detailed information regarding the methods has been previously published²⁴ and therefore is only summarized here. The Concussion Research Initiative of the Grand Alliance started in 2014 and involved NCAA varsity athletes and cheerleaders from 29 institutions who underwent yearly baseline and postconcussion assessments using standardized measures of concussion symptoms and clinical testing. The KD test was offered as an optional assessment (level B), and 5 institutions chose to include it in their protocols. The inclusion criteria were participation as a varsity NCAA intercollegiate student-athlete between July 2014 and August 2016 and a valid KD baseline score. The exclusion criteria were an invalid test (eg, incomplete test, errors on both tests) or incomplete participant information in the database. All participants provided written informed consent, as approved by each institutional review board and the University of Michigan Human Research Protection Office, at their respective institutions before the study.

Table 1. Participant Demographic and Anthropometric Values

Characteristic	No.	Males,		Age, y	Height, m	Weight, kg
		No.	%			
Between trials	3248	55.3		20.2 ± 2.3	1.78 ± 0.11	80.7 ± 21.0
Between years	833	59.8		19.8 ± .2	1.80 ± 0.11	82.1 ± 21.4

Procedures

Before sport activity, all participants completed a baseline KD assessment (King-DeVick Test, Inc, Oakbrook Terrace, IL) using either the spiral-bound cards (n = 2303) or tablet (n = 945) version as part of the CARE protocol and consistent with the manufacturer's recommendations.²⁵ The KD test consists of 4 cards (card version) or screens (tablet version). The first card or screen is for demonstration purposes, with 3 subsequent testing cards or screens. The spiral-bound cards have 8 rows of 5 single-digit numbers placed at random positions in each row. Vertical separation is 0.75 in (1.91 cm) on cards 1 and 2 and 0.25 in (0.64 cm) on card 3. In the tablet version, the same numbers are presented with consistent spacing on consecutive screens. Using the spiral-bound cards, participants are asked to read out loud the digits on the 3 testing cards from left to right, top to bottom, as quickly as possible without making errors. The total time to complete the 3 test cards is the first trial time. Two trials are administered, and the fastest time without any errors is recorded as the final baseline test score. A mistake that is not quickly corrected before the individual moves on to the next digit is considered an *error*. Cards are held at normal reading distance and corrections for nearsightedness are allowed; however, participants are not allowed to trace the numbers on the cards with their finger or hand. The researcher used a stopwatch to obtain the time required for the participant to read the digits on each card. For the tablet version, the participant started and stopped each trial by touching the tablet screen. Throughout the study, all participants were consistently tested using either the cards or a tablet. According to the CARE protocol, they were tested annually; however, test administrations may have varied between years, which is common in concussion management.^{5,6}

Table 2. Participants' Sports

Sport	No. (%)	
	Between Trials	Between Years
Football	858 (26.4)	261 (31.3)
Cross-country/track and field	550 (16.9)	147 (17.6)
Soccer	360 (11.1)	117 (14.0)
Rowing/crew	218 (6.7)	13 (1.6)
Basketball	211 (6.5)	60 (7.2)
Volleyball	136 (4.2)	47 (5.6)
Baseball	187 (5.8)	40 (4.8)
Lacrosse	154 (4.7)	24 (2.9)
Swimming	153 (4.7)	31 (3.7)
Softball	122 (3.8)	23 (2.8)
Gymnastics	74 (2.3)	21 (2.5)
Tennis	80 (2.5)	20 (2.4)
Golf	78 (2.4)	17 (2.0)
Field hockey	35 (1.1)	
Diving	25 (<1)	6 (<1)
Not reported	7 (<1)	6 (<1)

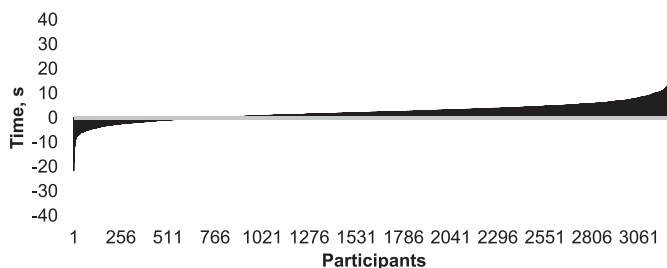


Figure 1. Ranked individual change between trials 1 and 2 for all participants (N = 3248). The time is calculated as trial 1 minus trial 2, such that a positive number reflects a decreased (faster) performance in trial 2.

Statistical Analysis

This was a prospective longitudinal study, and the dependent variable was the best total time (score) for each participant. Descriptive statistics were calculated for the KD test score, and differences were calculated as second administration minus first administration either within (aim 1) or between (aim 2) years; thus, negative numbers reflected faster performance. To address the 3 purposes of this study, agreement between measurements was analyzed using 2-way mixed-effects ICCs with absolute agreement (ICC [2,1]). For aim 1, the test-retest reliability between trials for all participants with 2 valid test scores during their year 1 baseline test (N = 3248) was compared. The ICC values were interpreted as <0.5 = *poor*, 0.5 to 0.74 = *moderate*, 0.75 to 0.90 = *good*, and >0.90 = *excellent*.²⁷ To evaluate the second aim, test-retest reliability over the course of a year, participants with 2 valid test scores a year apart (n = 833) were calculated. Finally, reliability was stratified by modality (cards = 566, tablet = 264) over the 2 years. We compared changes between trials and years using paired-samples *t* tests and calculated effect sizes (Cohen *d*).

RESULTS

Trial Test-Retest

A significant intraclass correlation was present within years between trials 1 and 2 for all participants (trial 1 = 43.2 ± 8.3 seconds, trial 2 = 40.8 ± 7.8 seconds; ICC = 0.888, *P* < .001). The paired-samples *t* test identified a change between trials (*t* = 34.9; 95% CI = 2.22, 2.48 seconds; *P* < .001, *d* = 0.29). The mean change between trials was -2.4 ± 3.8 seconds, the mode was 0.0 seconds, and the median was -2.2 seconds. Most participants (76.8%, 2495 of 3248) improved (ie, had a faster time) on trial 2 (Figure 1).

Annual Test-Retest

A significant positive correlation existed between year 1 and year 2 test scores across modalities (year 1 = 40.8 ± 7.4 seconds, year 2 = 38.7 ± 7.7 seconds; ICC = 0.827, *P* < .001). We noted a difference between the 2 years (*t* = 13.1; 95% CI = 1.71, 2.32 seconds; *P* < .001, *d* = 0.27). The mean change between years was -2.0 ± 4.5 seconds, the mode was -2.0 seconds, and the median was -2.1 seconds. Most participants (72.6%, 605 of 833) improved (had a faster time) during year 2 testing (Figure 2).

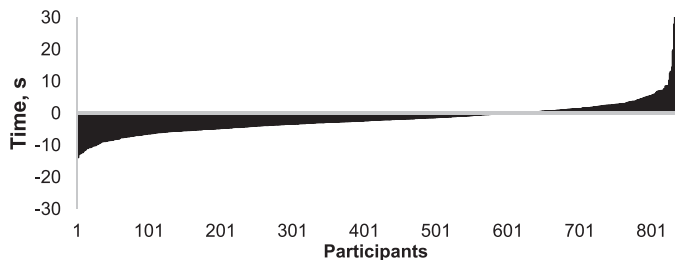


Figure 2. Ranked individual change between years 1 and 2 for all participants ($N = 833$). The time is calculated as year 1 minus year 2, such that a positive number reflects a decreased (faster) performance in year 2.

Stratified by Modality Reliability

When the tests were stratified by modality, the significant positive correlations for both versions persisted. For the spiral-bound cards, the ICC was 0.834 (year 1 = 39.9 ± 7.3 seconds, year 2 = 38.4 ± 8.1 seconds; $P < .001$), which represented a change ($t = 7.84$; 95% CI = 1.11, 1.85 seconds; $P < .001$, $d = 0.19$) between years. The mean change between years was -1.5 ± 4.5 seconds, the mode was -1.5 seconds, and the median was -1.8 seconds. Most participants (67.7%, 383 of 566) improved (had a faster time) during year 2 testing.

For the tablet version, a significant positive correlation was demonstrated between years 1 and 2 (year 1 = 42.7 ± 7.2 seconds, year 2 = 39.5 ± 6.8 seconds; ICC = 0.827, $P < .001$), which represented a change ($t = 7.84$; 95% CI = 2.67, 3.67 seconds; $P < .001$, $d = 0.46$). The mean change between years was -3.2 ± 4.1 seconds, the mode was -2.9 seconds, and the median was also -2.9 seconds. Most participants (84.1%, 222 of 264) improved (had a faster time) during year 2 testing.

DISCUSSION

To our knowledge, this was the first large-scale, independent study examining the reliability of the KD test across trials, between 2 years, and stratified by test modality among intercollegiate student-athletes. Our key findings suggest the KD test had good reliability (ICC = 0.75–0.90) between trials, between years, and when stratified by test modality, with a modest, though clinically relevant, learning effect over time.²⁷ Clinically, it is important to note that repeat administration annually resulted in mean improvements (faster performance) of -2.0 ± 4.5 seconds; 73% of student-athletes improved their performance from year 1 to year 2, which was associated with a small effect size. However, 27% of participants had poorer (slower) performance in year 2 (range = 0.01–31.2 seconds slower), raising concerns about test specificity. When we considered test modality, both the card (ICC = 0.834) and tablet (ICC = 0.827) versions had similar reliability. Taken together, these results suggest that the KD assessment was a reliable test with modest learning effects over time and that the assessment modality did not adversely affect baseline reliability.

The test manufacturer recommends 2 trials at baseline, and we found that test-retest performance between trials 1 and 2 during year 1 was reliable (ICC = 0.888). The overall mean time for all trials over both years was 40.4 ± 7.6 seconds, which was slightly slower but generally consistent

with prior KD results among collegiate student-athletes (36.3–38.7 seconds).^{14,15,28,29} It is worth noting that previous authors predominantly studied samples of 152 to 220 football and basketball players. The largest investigation consisted of 755 participants, whereas we assessed participants in 16 sports with more than 3200 baseline tests. The improvement from trial 1 to trial 2 during year 1 (-2.4 ± 3.8 seconds) was also generally consistent with earlier findings,¹⁴ further suggesting a modest learning effect. Although most participants (76.8%) improved on the second trial, a minority (22.4%, 725 of 3248) were slower (mean \pm SD = 2.2 ± 2.2 seconds, mode = 0.30 seconds, median = 1.6 seconds) on the second trial. The reasons for this decreased performance are unknown, but the finding reinforces the need to perform multiple trials at baseline. Future researchers should determine the number of trials required at baseline before the participant's performance truly plateaus and a stable outcome is identified. As with most concussion-related baseline assessments, attention must be paid to the motivation to minimize the effect of poor effort, and future investigators could assess both motivational strategies and effort level for achieving maximal performance. To date, no internal validity measure exists for the KD test; such measures are traditionally used for cognitive testing.³⁰

The KD test instructions recommend an annual baseline test for individuals over the age of 10. Our results indicated the annual test-retest performance displayed good reliability (ICC = 0.827), which was generally consistent with prior studies (ICC = 0.81–0.90) but higher than in the preliminary CARE study (ICC = 0.74).^{20,31–33} The overall improvement in baseline performance was -2.0 ± 4.5 seconds, which was less than in prior studies^{14,27,31,34} of smaller populations involving shorter intervals between tests, where improvements of 2.8 to 6.7 seconds were shown. Although the measures of central tendency demonstrated mild improvements, likely secondary to a practice or learning effect, the overall range of performance changes between years was 31.2 seconds (slower) to -16.4 seconds (faster). Furthermore, more than one-quarter (27.1%) of participants performed more slowly on the retest, with a mean increase (slower) of 3.2 ± 3.9 seconds. The manufacturer¹³ stated that any increased time (slower performance) represented a “failed” test,²⁵ and in individuals with suspected concussions, slower KD performance times were associated with a 5 times greater risk of concussion. Therefore, clinicians need to consider that more than one-quarter of collegiate student-athletes may perform more slowly during a subsequent test administration, independent of a concussion, which reinforces the concept that concussion is a clinical diagnosis supported by a multifaceted assessment battery and no single test should ever be used as a standalone diagnostic modality.^{2–4} Finally, given the time-consuming nature of retesting hundreds of student-athletes each year, overall test reliability, cost of the tablet-based assessment, and the 27.1% “failure” rate raise serious questions as to the necessity of annual KD baseline testing in collegiate student-athletes.^{35,36} An important future study will assess the sensitivity and specificity of the KD test acutely postconcussion in individuals who have had multiple baseline tests.

The manufacturer now recommends that only the electronic tablet version of the assessment be used because of reported higher test-retest reliability than for the spiral-bound cards, incompatibility across test platforms, and improved testing standardization and to minimize administration errors.^{23,25} We identified similar, but not equal, reliability between modalities, with slightly better reliability for the spiral-bound cards (ICC = 0.834) than the tablet (ICC = 0.827) version. Previously, Raynowska et al²³ demonstrated excellent agreement with a strong positive correlation (ICC = 0.92) between modalities (spiral-bound cards versus tablet) across a wide range of ages (7–25 years) when the same participants took both tests, but all test results were included, regardless of the number of errors committed. Conversely, our participants used the same modality each time, the tests were annual, and only error-free trials were considered as baseline scores, which could explain the lower reliability. From a reliability perspective, these results do not support the elimination of the spiral-bound cards for test administration. Further, the spiral-bound cards may be more clinically feasible given the substantial costs associated with the tablet version of the test. Among collegiate student-athletes, the spiral-bound cards resulted in faster (2.8 seconds) performance than the tablet version, but the reliability was not established.²² Thus, further investigation of postinjury specificity and sensitivity between modalities is required to more thoroughly address the modality selection.

These results are limited to collegiate student-athletes whose institutions were participating in the CARE study, which must be considered when extrapolating to other populations. The participants represented a diverse array of 16 sports at 5 institutions, which is a strength of the research. However, despite the standardized instructions, the tests were administered by numerous staff members at 5 institutions and were thus subject to intertester variability, which likely mirrors clinical practice.⁵ Moreover, to maintain the ecological validity of the outcomes, we did not exclude participants who sustained concussions during the year, which could have influenced the test outcomes. As the KD test has become more recognized in recent years, it is possible participants had taken this test before becoming involved in CARE (eg, in high school), which might have influenced the reliability outcomes. Finally, the test's reliability over additional years is currently unknown.

In this large sample of collegiate athletes across 16 sports, the KD test was reliable from trial to trial (ICC = 0.888), between years (ICC = 0.827), and when stratified by modality (spiral-bound cards ICC = 0.834, tablet ICC = 0.827). As expected, a mild practice effect was noted with a 2-second improvement between years, but caution must be exercised when interpreting the outcomes. Wide performance ranges were noted, with 27.1% of participants having slower KD times in year 2 than in year 1, which could be misclassified as “failing” the test, thereby reinforcing the importance of having a trained health care provider make the clinical diagnosis of concussion. Future independent investigators should continue to explore KD test determinants and psychometric properties, both at baseline and postconcussion, to achieve continuous improvements in concussion diagnosis and management.

ACKNOWLEDGMENTS

We acknowledge the research team members and clinical athletic trainers who assisted in the data collection for this project. This publication was made possible, in part, by support from the Grand Alliance CARE Consortium, funded by the NCAA and the DoD. The US Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick, MD 21702-5014 is the awarding and administering acquisition office. This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Psychological Health and Traumatic Brain Injury Program under award W81XWH-14-2-0151. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the DoD (Defense Health Program funds). King-Devick Technologies, Inc, provided iPad tablets to test sites free of charge as part of the overall study.

REFERENCES

1. Coronado VG, Haileyesus T, Cheng TA, et al. Trends in sports- and recreation-related traumatic brain injuries treated in US emergency departments: the National Electronic Injury Surveillance System-All Injury Program (NEISS-AIP) 2001–2012. *J Head Trauma Rehabil.* 2015;30(3):185–197.
2. Garcia GP, Lavieri MS, Jiang R, McAllister TW, McCrema MA, Broglio SP. A data-driven approach to unlikely, possible, probable, and definite acute concussion assessment. *J Neurotrauma.* 2019;36(10):1571–1583.
3. Garcia GP, Broglio SP, Lavieri MS, McCrema M, McAllister T; CARE Consortium Investigators. Quantifying the value of multidimensional assessment models for acute concussion: an analysis of data from the NCAA-DoD Care Consortium. *Sports Med.* 2018;48(7):1739–1749.
4. McCrory P, Meeuwisse W, Dvorak J, et al. Consensus statement on concussion in sport: the 5th International Conference on Concussion in Sport held in Berlin, October 2016. *Br J Sports Med.* 2017; 51(11):838–847.
5. Buckley TA, Burdette G, Kelly K. Concussion-management practice patterns of National Collegiate Athletic Association Division II and III athletic trainers: how the other half lives. *J Athl Train.* 2015;50(8):879–888.
6. Kelly KC, Jordan EM, Joyner AB, Burdette GT, Buckley TA. National Collegiate Athletic Association Division I athletic trainers' concussion-management practice patterns. *J Athl Train.* 2014;49(5):665–673.
7. Master CL, Scheiman M, Gallaway M, et al. Vision diagnoses are common after concussion in adolescents. *Clin Pediatr (Phila).* 2016;55(3):260–267.
8. Yorke AM, Smith L, Babcock M, Alsalaheen B. Validity and reliability of the Vestibular/Ocular Motor Screening and associations with common concussion screening tools. *Sports Health.* 2017;9(2):174–180.
9. Heitger MH, Jones RD, Anderson TJ. A new approach to predicting postconcussion syndrome after mild traumatic brain injury based upon eye movement function. *Conf Proc IEEE Eng Med Biol Soc.* 2008:3570–3573.
10. King A, Devick S. *The Proposed King-Devick Test and Its Relation to the Pierce Saccade Test and Reading Levels.* Chicago, IL: Illinois College of Optometry; 1976.
11. Galetta KM, Barrett J, Allen M, et al. The King-Devick test as a determinant of head trauma and concussion in boxers and MMA fighters. *Neurology.* 2011;76(17):1456–1462.
12. Ventura RE, Balcer LJ, Galetta SL. The neuro-ophthalmology of head trauma. *Lancet Neurol.* 2014;13(10):1006–1016.
13. Galetta KM, Morganroth J, Moehring N, et al. Adding vision to concussion testing: a prospective study of sideline testing in youth and collegiate athletes. *J Neuroophthalmol.* 2015;35(3):235–241.

14. Galetta KM, Liu M, Leong DF, Ventura RE, Galetta SL, Balcer LJ. The King-Devick Test of rapid number naming for concussion detection: meta-analysis and systematic review of the literature. *Concussion*. 2015;1(2):CNC8.
15. Marinides Z, Galetta KM, Andrews CN, et al. Vision testing is additive to the sideline assessment of sports-related concussion. *Neurol Clin Pract*. 2015;5(1):25–34.
16. Molloy JH, Murphy I, Gissane C. The King-Devick (K-D) test and concussion diagnosis in semi-professional rugby union players. *J Sci Med Sport*. 2017;20(8):708–711.
17. Burk JM, Munkasy BA, Joyner AB, Buckley TA. Balance Error Scoring System performance changes after a competitive athletic season. *Clin J Sport Med*. 2013;23(4):312–317.
18. McCrea M, Kelly JP, Randolph C, Cisler R, Berger L. Immediate neurocognitive effects of concussion. *Neurosurgery*. 2002;50(5):1032–1040.
19. Galetta MS, Galetta KM, McCrossin J, et al. Saccades and memory: baseline associations of the King-Devick and SCAT2 SAC tests in professional ice hockey players. *J Neurol Sci*. 2013;328(1–2):28–31.
20. Broglio SP, Katz BP, Zhao S, McCrea M, McAllister T; CARE Consortium Investigators. Test-retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD CARE Consortium. *Sports Med*. 2018;48(5):1255–1268.
21. Chrisman SPD, Harmon KG, Schmidt JD, et al. Impact of factors that affect reading skill level on King-Devick baseline performance time. *Ann Biomed Eng*. 2019;47(10):2122–2127.
22. Clugston JR, Chrisman SPD, Houck ZM, et al. King-Devick Test time varies by testing modality [published online ahead of print October 23, 2018]. *Clin J Sport Med*. doi:10.1097/JSM.0000000000000691.
23. Raynowska J, Hasanaj L, Zhang I, et al. Agreement of the spiral-bound and computerized tablet versions of the King-Devick Test of rapid number naming for sports related concussions. *Ann Sports Med Res*. 2015;2(9):1051–1057.
24. Broglio SP, McCrea M, McAllister T, et al. A national study on the effects of concussion in collegiate athletes and US military service academy members: the NCAA-DoD Concussion Assessment, Research and Education (CARE) Consortium Structure and Methods. *Sports Med*. 2017;47(7):1437–1451.
25. Frequently asked questions. King-Devick Technologies Web site. <https://kingdevicktest.com/concussions/faq/>. Accessed September 5, 2017.
26. Broglio SP, Katz BP, Zhao S, McCrea M, McAllister T; CARE Consortium Investigators. Test-retest reliability and interpretation of common concussion assessment tools: findings from the NCAA-DoD care consortium. *Sports Med*. 2018;48(5):1255–1268.
27. Portney L, Watkins M. *Foundations of Clinical Research: Applications to Practice*. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2000.
28. Galetta KM, Brandes LE, Maki K, et al. The King-Devick test and sports-related concussion: study of a rapid visual screening tool in a collegiate cohort. *J Neurol Sci*. 2011;309(1–2):34–39.
29. Leong DF, Balcer LJ, Galetta SL, et al. The King-Devick test for sideline concussion screening in collegiate football. *J Optom*. 2015;8(2):131–139.
30. Erdal K. Neuropsychological testing for sports-related concussion: how athletes can sandbag their baseline testing without detection. *Arch Clin Neuropsychol*. 2012;27(5):473–479.
31. Oberlander TJ, Olson BL, Weidauer L. Test-retest reliability of the King-Devick Test in an adolescent population. *J Athl Train*. 2017;52(5):439–445.
32. Alsalaheen B, Haines J, Yorke A, Diebold J. King-Devick Test reference values and associations with balance measures in high school American football players. *Scand J Med Sci Sports*. 2016;26(2):235–239.
33. Smolyansky V, Morettin CE, Hitzman SA, Beckerman S. Test-retest reliability of the King-Devick Test in elite junior Olympic athletes. *Optom Visual Perform*. 2016;4(3):147–154.
34. Heick JD, Bay RC, Dompier TP, Valovich McLeod T. The psychometric properties of the King-Devick Test and the influence of age and sex in healthy individuals aged 14 to 24 years. *Athl Train Sports Health Care*. 2016;8(5):222–229.
35. Buckley TA, Baugh CM, Meehan WP III, DiFabio MS. Concussion management plan compliance: a study of NCAA Power 5 Conference schools. *Orthop J Sports Med*. 2017;5(4):2325967117702696.
36. Katz BP, Kudela M, Harezlak J, McCrea M, McAllister T, Broglio SP. Baseline performance of NCAA athletes on a concussion assessment battery: a report from the CARE Consortium. *Sports Med*. 2018;48(8):1971–1985.

Address correspondence to Thomas A. Buckley, EdD, ATC, Department of Kinesiology and Applied Physiology, University of Delaware, 349 Tower at STAR Campus, 100 Discovery Boulevard, Newark, DE 19716. Address e-mail to TBuckley@UDel.edu.