On-field assessment of concussion: clinical utility of the King-Devick test

Andrew D Legarreta1, Nishit Mummareddy1, Aaron M Yengo-Kahn1,2, Scott L Zuckerman1,2
1Vanderbilt Sports Concussion Center, Vanderbilt University School of Medicine, Nashville, TN, USA; 2Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract: Sport-related concussion (SRC) is an important public health concern with up to 3.8 million SRCs occurring each year. As the incidence and rate of SRC increases, reliable and valid tools for diagnosis and management are needed. The King-Devick (K-D) test assesses a patient’s visual function based on reading a series of numbers as well as counting both time to completion and errors. Its rapid administration time and simplicity make the K-D test a potentially useful SRC diagnostic tool, though limitations exist in baseline testing and what constitutes an abnormal score. Additionally, the K-D tests should never be used in isolation to diagnose a concussion, but rather as one test in conjunction with additional clinical measures, as part of an individualized approach to each patient. The current review examines the clinical utility of the K-D test.

Keywords: King-Devick test, sport-related concussion, narrative review, SRC

Introduction

Sport-related concussion (SRC) and mild traumatic brain injury (mTBI) continue to represent an important public health problem for the scientific and general community alike. An estimated 3.89 SRCs per 10,000 athlete-exposures (AE) occur across all high-school athletes.1 Football had the highest rate with 9.21 SRCs per 10,000 AEs.1 As the burden of SRC weighs heavily on patients and health care providers, effective SRC diagnosis and management – from sideline, to clinic, to long-term management – represents an important area of investigation.2

Perhaps the most important step in SRC management is the initial diagnosis. A concussion diagnosis is often made on the field of play or shortly thereafter. While no gold-standard sideline diagnostic test currently exists, several screening exams are used in conjunction with individualized clinical decision making to arrive at a diagnosis. A frequently studied tool is the King-Devick Test®,3,4 in which a patient’s visual functions are assessed by reading a series of numbers and scored based on time to completion and number of errors. This paper will provide a narrative review of the King-Devick (K-D) test including its sideline use, with a focus on its utility to the initial diagnosis of an SRC.

Background

Poor oculomotor function is a reliable discriminator of mild traumatic brain injury (mTBI) due to impairment of both afferent and efferent visuomotor function.5,6 Afferent pathways include the optic nerve, white matter tracts, and visual cortices, while efferent pathways include cranial nerves responsible for normal eye
movements and vestibular functioning. While both visual systems can be affected after mTBI, injury to the efferent visual systems can lead to significant reading and concentration difficulties. Efferent or oculomotor-based dysfunctions involve version (e.g., saccade), vergence (e.g., convergence), and accommodation. Several reports suggest that oculomotor testing after mTBI is a useful, non-invasive means of clinical assessment. In particular, a study at the University of Pittsburgh of 64 patients and 78 controls undergoing the Vestibular/Ocular Motor Screening (VOMS) assessment demonstrated excellent internal consistency and high sensitivity when assessing ocular and vestibular dysfunction in concussed athletes.

The K-D test has been described as a means to assess saccadic eye movements, concentration, and attention. The K-D test was designed by Alan King and Steven Devick in the 1970s for evaluation of children suspected of dyslexia or impaired saccadic eye movements. Previously, it had also found use in studies of sleep-deprivation and reading ability in elementary school students. In recent years, its ease of use and potential for rapid administration in less than 2 minutes has made it a commonly used sideline tool for evaluation of suspected SRC. Research utilizing the K-D test has been performed in collegiate football players, collegiate soccer players, collegiate basketball players, amateur rugby players, professional ice hockey players, boxers, and mixed martial arts (MMA) fighters. The K-D test relies on a comparison of differences between baseline and post-injury results to provide insight into a potential brain injury.

The test
The K-D test utilizes numbers placed in horizontal lines for a total of three different patterns (Figure 1). After receiving standardized instructions and practicing with a demonstration card, patients are asked to read the sequence of numbers from left-to-right as quickly as possible. The total time required to read all three patterns is recorded in seconds using a stopwatch. Errors are recorded when a patient states the incorrect number without correcting the error.

The K-D test has potential clinical value when used as a tool to compare individualized baseline and post-injury values. Inferences about the occurrence of a possible SRC – not a diagnosis itself – are made when comparing post-injury, sideline testing to a pre-existing baseline measure. The rapid administration time and simplicity make it potentially suitable for sideline SRC evaluation by medical professionals, in conjunction with other validated measures. In order to simulate the noisy environment of an athletic field, baseline values for the K-D test have been obtained in noisy locker rooms, where the patient may have a more challenging time paying attention, although the test

**Figure 1** Sample King-Devick Test Card.
publisher does not explicitly require any particular baseline testing conditions. One unique aspect of the K-D test is its suggested administration by parents and non-health care professionals. A study of amateur boxers suggested that the K-D test can be accurately administered by non-medically trained parents or guardians with acceptable reliability.

Reliability, validity, and normative data

In order for an assessment tool to be clinically useful, it first needs to demonstrate reliability and subsequent validity. Reliability is a test’s capability of consistently producing the same result. Validity is the accuracy of a test’s results, typically measured against a well-established gold-standard. In the context of SRC, a reliable and valid assessment tool would consistently detect significant post-concussion changes.

Reliability of the K-D test has been evaluated in various populations. Galetta and colleagues found that the K-D test demonstrated high levels of test–retest reliability in boxers and MMA fighters both in successive tests before a fight and immediately pre- and post-fight, suggesting acceptable reliability in the short term. The authors reported an intraclass correlation coefficient (ICC) of 0.97 (95% confidence interval 0.90–1.0) between the two prefight sessions. In a separate study of college football players, test–retest reliability (ICC) was estimated at 0.95 (95% confidence interval 0.85–1.05) between baseline and post-season testing sessions of athletes. Interestingly, the high test–retest reliability of the K-D test remained when administered by non-professional parents. Another study by Hecimovich and colleagues examined Australian football players and reported an ICC of 0.91 (CI 0.80–0.96) between two baseline testing sessions, with a range of 1–12 weeks between these tests. Conversely, in a study of 68 healthy adolescents, ICC was 0.81 (CI 0.73–0.87) with a high repeatability coefficient (±8.76 s), which represents 95% of the differences between repeated measures of the same participants.

The authors suggested a large learning effect may limit the utility of the K-D test, despite a clinically acceptable ICC. A recent study evaluated the test–retest reliability of the K-D test at 3 time points (baseline, pre-practice, and removal from practice) in 45 healthy high school students. This study found that the K-D test had an overall reliability as measured by ICC of 0.91 [95% CI, 0.86–0.95], baseline to pre-practice of 0.90 [95% CI, 0.82–0.94], and baseline to removal from practice of 0.87 [95% CI, 0.76–0.93].

Several studies have aimed to determine the validity of the K-D test. In 337 high school football players from four Southwest Michigan high schools, K-D test times of all 9 concussed football players were worse than their baseline times (66.2 s sideline vs 47.2 s baseline, p=0.001). The remaining non-concussed athletes had only minimal, non-statistically significant changes relative to their baseline. In a study of 219 male and female collegiate athletes at the University of Pennsylvania, baseline K-D testing was performed prior to the start of the 2010–11 season in athletes participating in the sports of football, soccer, and basketball. The 10 patients who sustained an SRC demonstrated significant worsening in their immediate sideline K-D test scores versus baseline, 46.9 s vs 37.0 s, p=0.009, respectively. Marinides and colleagues investigated the use of the K-D test as a complement to the Standardized Assessment of Concussion (SAC) and Balance Error Scoring System (BESS) in a cohort of University of Florida men’s football, women’s soccer, and women’s lacrosse teams. Of the total 217 athletes assessed at baseline, 29 concussed patients underwent post-injury testing. Most patients (23/28) had testing done on the same day of injury, while the remaining (5/28) had testing done in the following days after their injury. Among K-D test scores, 23/29 (79%) concussed patients had worse performance from baseline to post-injury; however, it should be noted that 21% exhibited no significant changes. The authors estimated that for every 1-point of worsening of overall SAC score, there was a 4.7-s increase in K-D total time score from baseline (p=0.001), while every 1-point worsening of BESS had an estimated 1.6-point increase in K-D time score from baseline (p=0.043). Lawrence and colleagues studied the diagnostic and prognostic value of the K-D test in an outpatient setting. They administered the K-D test pre- and post-treadmill test to both individuals who had a concussion recently and baseline, 46.9 s vs 37.0 s, p=0.009, respectively. The remaining non-concussed athletes had only minimal, non-statistically significant changes relative to their baseline. The authors suggested a large learning effect may limit the utility of the K-D test, despite a clinically acceptable ICC. A recent study evaluated the test–retest reliability of the K-D test at 3 time points (baseline, pre-practice, and removal from practice) in 45 healthy high school students. This study found that the K-D test had an overall reliability as measured by ICC of 0.91 [95% CI, 0.86–0.95], baseline to pre-practice of 0.90 [95% CI, 0.82–0.94], and baseline to removal from practice of 0.87 [95% CI, 0.76–0.93].

Several studies have aimed to determine the validity of the K-D test. In 337 high school football players from four Southwest Michigan high schools, K-D test times of all 9 concussed football players were worse than their baseline times (66.2 s sideline vs 47.2 s baseline, p=0.001). The remaining non-concussed athletes had only minimal, non-statistically significant changes relative to their baseline. In a study of 219 male and female collegiate athletes at the University of Pennsylvania, baseline K-D testing was performed prior to the start of the 2010–11 season in athletes participating in the sports of football, soccer, and basketball. The 10 patients who sustained an SRC demonstrated significant worsening in their immediate sideline K-D test scores versus baseline, 46.9 s vs 37.0 s, p=0.009, respectively. Marinides and colleagues investigated the use of the K-D test as a complement to the Standardized Assessment of Concussion (SAC) and Balance Error Scoring System (BESS) in a cohort of University of Florida men’s football, women’s soccer, and women’s lacrosse teams. Of the total 217 athletes assessed at baseline, 29 concussed patients underwent post-injury testing. Most patients (23/28) had testing done on the same day of injury, while the remaining (5/28) had testing done in the following days after their injury. Among K-D test scores, 23/29 (79%) concussed patients had worse performance from baseline to post-injury; however, it should be noted that 21% exhibited no significant changes. The authors estimated that for every 1-point of worsening of overall SAC score, there was a 4.7-s increase in K-D total time score from baseline (p=0.001), while every 1-point worsening of BESS had an estimated 1.6-point increase in K-D time score from baseline (p=0.043). Lawrence and colleagues studied the diagnostic and prognostic value of the K-D test in an outpatient setting. They administered the K-D test pre- and post-treadmill test to both individuals who had a concussion recently and matched controls, and found that concussed individuals had significantly higher abnormal repetitive saccades than matched controls. Furthermore, they reported that individuals with concussions who required more than 2 weeks to recover did not improve on their K-D test after the treadmill performance whereas both matched controls and concussed individuals who required less than 2 weeks to recover did.
Outside the US, in a study evaluating the validity of the K-D test in Australian football players, the authors reported an overall specificity and sensitivity of 0.96 and 0.98 for the diagnosis of SRC, respectively. Concussed patients had a 2-s worsening in their post-concussive K-D scores \((p<0.0001)\), with the medical SRC diagnosis being made by a comprehensive history and SCAT3 performed by a health care provider.\(^8\) First, this 2-s change on the K-D is starkly different than similar studies that report changes of 8–10 s. Second, it is also unclear if the K-D result included only worsening time needed to read the numbers, the number of errors committed, or both.\(^9\)

The lack of a K-D threshold – time or errors – in this study limits any major conclusions from being made. It is essential to note; however, that SRC diagnosis was made by a health care professional using the SCAT3 and not the K-D test in isolation. Lastly, Fuller and colleagues reported significantly lower specificity and sensitivity values of 0.39 and 0.60, respectively, in a cohort of 145 rugby players.\(^10\) In the administration of the K-D test, any worsening of time and/or errors committed from the baseline values was considered an abnormal result.\(^11\) A concussion diagnosis was made using the Head Injury Assessment (HIA) test, performed by a team doctor.\(^12\) Overall, the data is mixed for K-D validity and inconsistencies exist among chosen threshold values for the determination of abnormal K-D test results. In many instances, the criteria of an abnormal K-D test result are unclear, making application of these study results difficult.

**Normative and baseline values**

Establishing normative baseline values is important for a clinical assessment tool so that injured patients without a baseline assessment have normative values for comparison. Whereas many neurocognitive and balance test measures used in SRC assessment have large normative standardization samples,\(^11\) a paucity of data exists with the K-D test, thus limiting its utility.\(^12\) One study of 422 young student-athletes between the ages of 8 and 14 found that age and gender significantly affected baseline test results; however, a history of concussion did not.\(^13\) Normative values were not provided in the study. Conversely, a study assessing baseline K-D values for male ice hockey players found no association with age or self-reported concussion history on K-D test performance.\(^14\) Alsalaheen and colleagues studied 157 high school football players between the ages of 13 and 18 years without a concussion in the prior 3 months to establish normative values in this population.\(^15\) In this cohort, participants aged 16–18 years performed better (median=40.9 s) on the K-D test compared to those aged 13–15 years (median=43.81 s).\(^16\) Furthermore, they reported that there was no difference in performance between individuals with a concussion history and those without a concussion history \((p=0.44)\).\(^17\) A small study of 12 healthy individuals between the ages of 24 and 49 reported an average reading time of 51.24 s and average number of saccades per person to be 145.\(^18\)

As stated earlier, normative values, such as those which have been determined and published for SRC tools such as balance testing and neurocognitive assessment, are crucial to determining clinical deviations from normal function in the absence of baseline scores.\(^19\) Because the K-D test relies on relative changes in values rather than absolute changes in values, establishing the reference values by which the test derives meaning is a critical aspect of the test’s clinical utility. Thus, it is imperative that either normative, baseline, or both types of values are available for the interpreting clinician. Several studies have begun investigating and reporting normative values for the K-D test, but more studies with larger populations incorporating additional potential contributory factors (gender, sport-type, weight, learning effect) need to be performed to establish robust normative data.

**Correlation with other measures**

In 27 professional ice hockey athletes who underwent pre-season modified SCAT2 and K-D testing, K-D test scores were found to be associated with reductions in immediate memory (as measured using the SCAT2 SAC immediate memory score).\(^20\) The authors postulated that these effects were explained by overlap of anatomic structures controlling working memory and saccadic eye movements, notably the dorsolateral prefrontal cortex.\(^20\) Similarly, K-D test scores have been found to be correlated with Military Acute Concussion Evaluation (MACE) scores.\(^21\) However, in a study of 105 healthy adolescents without a history of concussion over the past 3 months, the authors did not observe any statistically significant relationships between the VOMS items and the K-D or BESS scores.\(^22\) The authors postulated that the three tests utilized were measuring different components of the ocular and vestibular systems, thus leading to poor correlation.\(^22\) Similarly, a study of high school football players found no correlation between the K-D test and the BESS \((p=0.86)\) as well as the reaction time \((p=0.18)\) and directional control.
components of the Limits of Stability (LOS) test.27

In summary, the K-D test has demonstrated some correlative ability with SCAT2 and MACE but has not done so with VOMS or BESS. Future study is required to comprehensively assess the K-D test comparatively to other neurocognitive measures in both the concussed and non-concussed sports populations.

Clinical and suggested use

Although the K-D test has been investigated as a means of sideline concussion evaluation and has entered into a licensing agreement with the Mayo Clinic,33 it fundamentally remains just one of the many useful tools to aid in concussion diagnosis, rather than provide an independent, gold-standard assessment. Its reliance on differences between baseline and post-injury scores can help clinicians identify concussed patients. However, areas in need of improvement exist (Table 1). First, even with a baseline test, a threshold to determine an “abnormal” test remains unknown, as the test publisher states “any worsening (increased) K-D test time compared to the subject’s baseline” is a failure of the K-D Test.4 Previous studies have also failed to identify time and/or error thresholds for the determination of meaningful clinical change. Second, in the case of a patient with no individual baseline score, there exists a need to establish norms that account for age and/or gender variables, as part of an individualized approach to SRC diagnosis.34 Third, any thresholds that are identified must reconcile the differences between relative changes rather than absolute changes.

In recognition of its ease of use and potential utility in SRC diagnosis, we suggest several clinical practices to aid with its implementation (Box 1). First, if the K-D is to be used, baseline values should be obtained for all athletes before participation in sport begins. These baseline scores include two measures: total time for completion and number of errors. Ideally, the K-D baseline values should be assessed in an environment similar to that in which a post-concussion evaluation would occur. That is, a noisy, mildly distracting environment, such as a locker room, would be more representative of the environment in which a sideline post-concussion evaluation would take place as compared to a quiet, clinical exam room.10 Second, the facilitator of the test should have basic training in the components of the test. At this moment, we caution strongly against using the K-D test by non-medical professionals for fear of relying solely on K-D results and possibly missing an injury, prematurely returning an athlete to play, or incorrectly identifying an SRC. Third, it is important to remember that the K-D test results do not preclude a comprehensive concussion evaluation and

### Table 1 Suggested areas of improvement

<table>
<thead>
<tr>
<th>Area of improvement</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish absolute threshold for abnormal test result (without baseline)</td>
<td>● Clear identification of what constitutes an abnormal test result is critical for clinical practicality of the K-D test ● Although the manufacturer accepts “any worsening” as an abnormal test result, this may exaggerate the sensitivity of the K-D test</td>
</tr>
<tr>
<td>Establish normative values</td>
<td>● In athletes without baseline testing, normative values for the general population would allow for substituted comparison values to assess changes after SRC ● Normative values for the general population would also allow for greater insight into K-D test reliability and validity</td>
</tr>
<tr>
<td>Reconciliation of relative changes with clinical recovery</td>
<td>● Understanding the clinical significance of relative value changes between two testing sessions could potentially help inform clinical progression of an SRC ● Correlate relative value changes with days to symptom recovery to improve utility of K-D test</td>
</tr>
</tbody>
</table>

**Abbreviations:** K-D test, King-Devick test; SRC, sport-related concussion.

**Box 1 Suggested clinical practices utilizing K-D test**

1. Obtain baseline K-D test values before sport participation, including total time for completion and number of errors generated.
2. Administer K-D only by medical professionals formally trained in all aspects of the K-D test and SRC management.
3. Provide comprehensive concussion evaluation to all athletes regardless of K-D test results.

**Abbreviations:** K-D test, King-Devick test; SRC, sport-related concussion.
should never replace the expertise of an experienced sports medicine professional. It must be emphasized that the K-D test evaluates one component – oculomotor function – of possible SRC signs and symptoms. In the absence of concrete, objective signs such as loss of consciousness, posturing, or seizures – each SRC diagnosis requires an individualized approach.

Summary
The K-D test is a straightforward, potentially useful supplementary concussion assessment tool. Though many studies have investigated the reliability and validity of the K-D test in a variety of sporting populations, the chosen study designs are inconsistent and there is a lack of well-established normative values from methodologically rigorous studies. Larger cohort studies examining unique demographic variables to ascertain accurate normative values, as well as thresholds for abnormal results, are needed.

Furthermore, the K-D test can be used as an adjunct to other concussion assessment tools such as the Sports Concussion Assessment Tool (SCAT), Balance Error Scoring System (BESS), and Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT®). Moreover, the K-D test should never be used in isolation and should only be employed by medical professionals. Areas most in need of further study include 1) pre- to post-injury change scores indicating potential clinical significance and 2) normative values for those without a baseline assessment.

Disclosure
The authors report no conflicts of interest in this work.

References


