Pocket Model, Numerical Readout Breath Alcohol Measurement Devices: A Laboratory- and In-vivo-Based Evaluation

WE Van Tassel¹
ME Dennis
ME Parker

¹Texas A&M University Center for Alcohol and Drug Education Studies, PO Box S-5, College Station, TX 77844, USA

Background
Breath alcohol concentration (BrAC), obtained most often for law enforcement forensic purposes, has been utilized internationally to combat alcohol impaired driving. Several companies currently offer hand-held, numerical-readout breath alcohol measurement devices, designed to provide civilians with information regarding their individual BrAC. The implied purpose of these prevention-oriented devices is to help drinkers make better decisions after consuming alcohol (i.e. whether to operate a motor vehicle).

Evidential breath testers (EBTs), preliminary breath testers (PBTs) and passive alcohol sensors (PASs) have all undergone rigorous laboratory and field-based analyses to evaluate their performance (1,2,3). However, there is a relative lack of evaluation of the lower-cost pocket model breath testers (PMBTs).

Objectives
The purpose of this exploratory study was to evaluate the performance of commonly available, quantitative-readout pocket model breath test devices.

Methodology
Eight (8) small scale, reusable breath alcohol testing devices were procured through internet-based and local retailers. Per device costs ranged from $40-100 USD. These PMBTs provided numerical readouts of BrAC, to the hundredth of one percent.

Experiment One
Experiment One assessed the devices at multiple alcohol levels, under laboratory conditions. The PMBTs were tested at six alcohol levels: .02, .04, .06, .08, .10 and .16%. Each device was tested 20 times at each alcohol level, resulting in a total of 120 tests per device. National Draeger Mark IIA simulators provided the alcoholic samples for this experiment (4).

Experiment Two
In Experiment Two, 10 volunteer participants (5M, 5F) agreed to consume alcohol and provide breath samples. Eligible participants were screened to exclude pregnant females, non-drinkers, heavy/problem drinkers, alcoholics, diabetics and those allergic to alcohol or in poor health.

Participants, after receiving training to provide consistent deep-lung breath samples, were dosed based on weight using alcoholic beverages consisting of 100 proof vodka and orange juice. A 15 minute deprivation period was observed prior to breath testing. Four BrAC testing levels were used: .08, .06, .04 and .02%.
An Intoxilyzer 5000 was used as the standard against which PMBT results were compared. When a participant was confirmed to be within a target testing range (.08, .06, .04, .02% ± .005), he or she provided samples with the Intoxilyzer and all test devices. Duplicate samples, which have been rated satisfactorily adequate for forensic purposes, were collected at every test, on every device (5).

Results and Analysis
Two devices failed to complete all testing, Device G displaying its maximum value (.19) on all tests and Device H ceasing to yield readings during Experiment Two. Because full and useful data were not obtained for these two devices, they were eliminated from further analysis. The six remaining instruments yielded complete data for all tests.

Experiment One
A total of 960 tests were performed in Experiment One. Table 1 shows the means and SDs of each device at each test level. Figure 1 graphically represents the mean results.

Table 1
Laboratory test results- means and standard deviations

<table>
<thead>
<tr>
<th>Device</th>
<th>.02</th>
<th>.04</th>
<th>.06</th>
<th>.08</th>
<th>.10</th>
<th>.16</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.010 (.000)</td>
<td>0.047 (.005)</td>
<td>0.094 (.005)</td>
<td>0.080 (.000)</td>
<td>0.095 (.005)</td>
<td>0.141 (.007)</td>
</tr>
<tr>
<td>B</td>
<td>0.036 (.005)</td>
<td>0.069 (.004)</td>
<td>0.117 (.007)</td>
<td>0.152 (.005)</td>
<td>0.196 (.008)</td>
<td>0.327 (.011)</td>
</tr>
<tr>
<td>C</td>
<td>0.030 (.000)</td>
<td>0.038 (.004)</td>
<td>0.082 (.009)</td>
<td>0.119 (.005)</td>
<td>0.155 (.008)</td>
<td>0.250 (.013)</td>
</tr>
<tr>
<td>D</td>
<td>0.020 (.000)</td>
<td>0.048 (.005)</td>
<td>0.084 (.009)</td>
<td>0.132 (.014)</td>
<td>0.181 (.019)</td>
<td>0.190 (.000)</td>
</tr>
<tr>
<td>E</td>
<td>0.014 (.005)</td>
<td>0.062 (.005)</td>
<td>0.060 (.000)</td>
<td>0.081 (.003)</td>
<td>0.109 (.003)</td>
<td>0.169 (.006)</td>
</tr>
<tr>
<td>F</td>
<td>0.034 (.005)</td>
<td>0.047 (.004)</td>
<td>0.057 (.004)</td>
<td>0.073 (.006)</td>
<td>0.080 (.007)</td>
<td>0.117 (.005)</td>
</tr>
</tbody>
</table>

Experiment Two
A total of 480 tests were performed during Experiment Two. Table 2 shows the means and SDs of each device at each test level. Figure 2 graphically shows the mean results, as compared to the Intoxilyzer. The mean results show that five out of the six devices read higher than the Intoxilyzer.

Repeated Measures Analysis of Variance revealed significant Main Effects for Device (F=29.667, p=.000) and for Concentration (F=147.270, p=.000) and a significant interaction between Device and Concentration (F=7.776, p=.000). Subsequent analysis using Simple Main Effects and Dunnett’s Test showed that only Device B was significantly different from the Intoxilyzer, only at the .08 test level.
Figure 1
Mean laboratory accuracy test results for each device at each simulator concentration.

Table 2
In-vivo test results- means and standard deviations

<table>
<thead>
<tr>
<th>Device</th>
<th>.02</th>
<th>.04</th>
<th>.06</th>
<th>.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intox</td>
<td>.023 (.001)</td>
<td>.039 (.005)</td>
<td>.058 (.004)</td>
<td>.080 (.004)</td>
</tr>
<tr>
<td>A</td>
<td>.060 (.016)</td>
<td>.083 (.018)</td>
<td>.104 (.018)</td>
<td>.152 (.042)</td>
</tr>
<tr>
<td>B</td>
<td>.069 (.021)</td>
<td>.122 (.039)</td>
<td>.146 (.046)</td>
<td>.178 (.059)</td>
</tr>
<tr>
<td>C</td>
<td>.038 (.008)</td>
<td>.070 (.012)</td>
<td>.095 (.022)</td>
<td>.123 (.018)</td>
</tr>
<tr>
<td>D</td>
<td>.042 (.012)</td>
<td>.073 (.034)</td>
<td>.096 (.033)</td>
<td>.093 (.040)</td>
</tr>
<tr>
<td>E</td>
<td>.038 (.007)</td>
<td>.065 (.015)</td>
<td>.081 (.013)</td>
<td>.103 (.012)</td>
</tr>
<tr>
<td>F</td>
<td>.020 (.009)</td>
<td>.033 (.017)</td>
<td>.037 (.012)</td>
<td>.043 (.010)</td>
</tr>
</tbody>
</table>
Figure 2
Mean In-vivo Intoxilyzer (dashed line) and device accuracy results for each device at each concentration. Test level is shown on X-axis, test result on Y-axis.

Discussion
For a breath alcohol testing device to be included on the National Highway Traffic Safety Administration’s (NHTSA) Conforming Products List, it must meet certain requirements regarding accuracy and precision (6). A device’s precision is acceptable if it demonstrates a Standard deviation of $\leq 0.0042$. Its accuracy is acceptable if it demonstrates a mean systematic error of $\pm 0.005$ BAC.

In Experiment One, no single device met the NHTSA criteria for precision or accuracy at all concentrations. The mean SD of the devices exceeded the NHTSA precision criteria by 31%. The PMBTs also exceeded the NHTSA accuracy criteria.

In Experiment Two, again no single device met the NHTSA criteria for precision or accuracy at all concentrations. The mean SD of the devices exceeded the NHTSA precision criteria by over 500%. The reduced precision results observed could reflect a decrease in consistency of breath samples from participants, compared to the higher level of consistency when tested under laboratory conditions.

The test devices, as a whole, tended to yield readings consistently higher than the Intoxilyzer readings. The sole exception to this finding was Device F, which consistently read below the Intoxilyzer results. One device, Device B, yielded readings up to 225% higher than the Intoxilyzer.
In terms of application, if these devices were to err, it could be argued that it would be preferred that they overestimate BAC. That is, it would be better if they yielded results higher than users’ actual BrACs, rather than lower than actual BrACs. An above-actual BrAC trend would be in what could be considered the “conservative” direction. If users employed the information from these devices to make decisions after consuming alcohol, it would be preferable if they believed their alcohol levels were at or higher than their actual levels (7).

Device F’s performance is of special concern in this respect. Its readings could cause a user to believe his or her BAC is lower than actual. If the user is employing the information to make a decision whether to operate a motor vehicle, the user might come to believe that his or her ability to drive is not impaired when in reality it is.

Whereas most other breath alcohol test devices are used for law enforcement and forensic purposes, PMBTs are designed for more casual use by civilians to assess their individual alcohol level. As such, PMBTs might not have to meet stringent performance criteria to be considered effective. It is possible that PMBTs could exhibit lower levels of precision and accuracy than other devices and still be considered fit for purpose.

Conclusion
The findings of this study suggest several conclusions. First, the PMBTs tested were successful in detecting the presence of alcohol. No false negatives or false positive readings were reported. Second, the PMBTs tended to read higher than actual BrAC, exceeding the standards observed using the simulators and the Intoxilyzer.

Third, the PMBTs’ performance was dependent upon the alcohol level at which they were tested. That is, the interaction between the devices and the alcohol test levels suggests that the PMBTs did not perform in a linear manner. As the alcohol level increased, the dependence upon the concentration levels became more variable. This could have implications regarding the devices’ use in actual drinking conditions, where some drinkers’ actual alcohol levels could be substantially above the maximum tested in Experiment Two.

Fourth, devices A, C and E were the best performing devices, in terms of accuracy and precision. Any of these three could have the potential to provide users with information that could be effectively used to make improved decisions regarding personal alcohol levels. The lower performing devices (B, D and F) would not be recommended for use by people seeking to obtain useful information regarding personal alcohol levels. Device F could be especially risky to use, as it consistently underestimated BrAC. However, more research is needed before any device receives solid recommendation or condemnation for use as intended.

Possible Next Steps
Several areas for future research are suggested by this study. First, any breath test device requires a waiting period to allow mouth alcohol and other potential contaminants to dissipate before testing. Future research could determine whether PMBT users would actually cease all alcohol and food consumption and wait a specified period of time before using a PMBT.

Second, PMBT users may have difficulty correctly operating the devices in an alcohol-positive state. Additional testing under actual drinking conditions could yield information regarding the realistic use of PMBTs. Third, users PMBT users might be unable to accurately interpret the devices’ readouts in an alcohol positive state. Future research
could assess this issue and compare the effectiveness of numerical versus categorical readout features.

Finally, the PMBTs were calibrated by their manufacturers. Only one PMBT tested permitted recalibration by the user. Additional research could reveal how long PMBTs’ calibration remains consistent, assessing device longevity, in terms of time and number of tests.

This exploratory study represents a step toward understanding the value of PMBTs in aiding drinkers to make improved decisions after consuming alcohol. Additional research could further reveal PMBTs’ role in helping to prevent alcohol impaired driving.

References