### Multimedia Appendix 4: Summary of the technical results of the clinician-centered studies

<table>
<thead>
<tr>
<th>Source (Health Condition)</th>
<th>Technical Results</th>
</tr>
</thead>
</table>
| Gillis et al, 2015 (Disaster relief) | Heat experiments:  
- For less computationally intensive applications, GG stabilizes around 95°F, but increases when running a video application  
- When running non-display background tasks (i.e. Wi-Fi connectivity) the heat stabilizes to around 73°F  

Range experiments:  
- Quality decreases while lag time increases with distance, but at large distances the quality is unusable and the device communication gets disconnected  

Average ease of setup in lake simulation: average score (1 = low to 5 = high)  
- Training on setup and use: 4.67  
- Assembly: 4.00  
- Equipment mobility: 4.00  

Average ease of use in lake simulation: average score (1 = low to 5 = high)  
- Video: 5.00  
- Audio: 2.67  
- Improved ability to serve purpose: 3.67  
- Equipment mobility: 4.00  
- Device and application: 4.00  
- Application stability: 3.33 |
| Cicero et al, 2014 (Disaster relief) | There was no significant difference in the triage accuracy between the intervention and control groups:  
- 100% of patients mistriaged by the intervention group were undertriaged, while 28.6% of the patients mistriaged for the control group were overtriaged  

The two patients triaged by the intervention group with help from the telemedicine physician took significantly more time to triage. |
| Newaz and Eide, 2015 (Disaster relief) | Time taken to navigate the route for participants using GG:  
- G1: 26.23 s  
- G2: 22.03 s  
- G3: 24.37 s  
- G4: 23.44 s  
- G5: 24.39 s  
- G6: 25.09 s  

Time taken to navigate the route for participants using mobile phones:  
- M1: 22.54 s  
- M2: 25.50 s  
- M3: 26.22 s  
- M4: 23.27 s  
- M5: 24.56 s |
<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
</table>
| Paxton et al, 2015 (Behavioral sciences) | Increases in lag significantly predicted decreases in r, the measure of interpersonal synchrony or convergence ($\beta = -0.50$, $p < 0.0001$):  
- Partners’ head movements were most strongly correlated at lag 0, or in moment-to-moment comparisons ($\beta = 0.19$, $p > 0.30$)  
There was no significant difference between the noise and dual-task conditions and no significant effect of the interaction term ($\beta = -0.03$, $p > 0.60$). |
| Pascale et al, 2015 (Nursing – peripheral detection) | Comparing monitor and GG:  
- Using GG, but leaving all other variables constant, significantly reduced the odds of detecting the dark gray tilted stimuli ($p < 0.001$)  
- Odds of detection also dropped significantly ($p < 0.001$) when stimuli were presented at the far eccentricity compared to near eccentricity  
- Presenting peripheral stimuli on GG at the far eccentricity dropped the odds of detection further ($p < 0.001$) |
| Yuan et al, 2015 (Neurology) | The teleneurohospitalist was able to quickly guide the local physician through a thorough evaluation and recommend the use of IV thrombolysis. The dysphasia resolved, the myodynamia of the right upper and lower limbs recovered from grade III to IV, and neurological function completely recovered, although an infarction spot was noted on CT scans. |
| Iversen et al, 2015 (Student training – physiotherapy students) | All students successfully completed the 160-minute session of vestibular dysfunction assessment and treatment skills:  
- Students in the GG group scored better on the clinical skills check, with a median score of 19 (range 16-20) versus 18 (range 16-20; $p = 0.03$)  
- There was no statistical difference between groups except in the category assessing execution and handling of patients during clinical assessment, in which 77.0% of the GG group and 59% of the control group performed satisfactorily ($p = 0.049$)  
- 37% of the GG group and 31% of the control group performed the execution and handling of the treatment manoeuver satisfactorily in every section of the test ($p = 0.58$) |
| Son et al, 2015 (Student training – otolaryngology residents) | Patient survey:  
- The scores of questions regarding the subject of better explanations ($p = 0.061$), listening carefully ($p = 0.123$), addressing patient questions ($p = 0.086$), spending adequate time ($p = 0.2$), and overall performance ($p = 0.0715$) decreased with the use of GG but none reached significance  
- The score for displaying respect decreased significantly ($p = 0.0065$) with the use of GG  
Faculty survey:  
- The scores of questions regarding making eye contact ($p < 0.001$), asking open-ended questions ($p = 0.0015$), sitting down ($p = 0.0095$), better explanations ($p < 0.001$), listening carefully ($p < 0.001$), addressing patient questions ($p < 0.001$), displaying respect ($p < 0.001$), and spending adequate time ($p = 0.0005$) significantly improved with use of GG. |
| GG | The scores for engaging with proper empathy (p = 0.0005) and overall performance (p = 0.014) significantly improved with the use of GG
| The score for proper instruction decreased with the use of GG but not to significance (p = 0.06)
| The score for handwashing increased with the use of GG but not to significance (p = 0.062) |

| Spaedy et al, 2016 (Radiology) | Average chest X-ray reading score (max. 23):
| Viewed through GG: 14.1 ± 2.2
| GG photo on mobile device: 18.5 ± 1.5
| Original X-ray on desktop computer: 21.3 ± 1.7
| (p < 0.0001 between GG and desktop computer, and p = 0.0004 between mobile device and desktop computer) |

| Russel et al, 2014 (Student training – medical students (radiology)) | Both telementored students and live instruction students were able to obtain adequate images for EPSS measurement in each instance:
| Telementored students had a median image quality rating of 7.5 out of 10 (IQR = 6 to 10)
| Live instruction students had a median image quality rating of 8 out of 10 (IQR = 7 to 9)
| Only one of the six subjects who received no instruction was able obtain an adequate image, and median quality ratings for this group were 0 out of 10 (IQR = 0 to 2)
| There was no significant difference in the achievement of adequate images for EPSS measurement, nor in image quality between telementored and live instruction groups |

| Wu et al, 2014 (Student training – medical students and radiology residents) | Central line performance metrics:
| The GG on average took longer to perform the procedure at every training level, but only reached significance for the M4 (197 s vs. 91 s, p ≤ 0.05) and PGY3 (151 s vs. 52 s, p ≤ 0.05) students
| In the PGY3 training level, participants spent significantly more time focusing their gaze on the patients (48 s vs. 23s, p ≤ 0.05), as well as on the GG monitor (103 s vs. 29 s, p ≤ 0.05) than the non-GG group
| MS1 students wearing GG spent significantly more time (139 s vs. 47 s, p ≤ 0.05) looking at the monitor than the non-GG group
| Both groups at all training levels had similar experiences (p > 0.05) performing central venous lines on both live and simulated patients |

| Widmer et al, 2014 (Dermatology and Radiology) | Accuracy when searching for skin of forearm image:
| Without a keyword, 3 of 10 images were accurate
| With a keyword, 7 of 10 images were accurate

Accuracy when searching for lung CT image: number of accurate results:
| Original photo (no keyword): 2/10
| Original photo (keywords lung+CT): 10/10
| Printed photo (no keyword): 0/10
| Printed photo (keywords lung+CT): 10/10
| LCD screen photo (no keyword): 1/10
<p>| LCD screen photo (keywords lung+CT): 7/10 |</p>
<table>
<thead>
<tr>
<th>Study (Author, Year)</th>
<th>Description</th>
<th>Measures/Benninger, 2015 (Radiology)</th>
<th>Class success range (ability to wear GG while using SonicEye ultrasound finger probe to identify a structure/space during a 1 minute laboratory examination:</th>
</tr>
</thead>
</table>
| Accuracy when searching for brain MRI image: number of accurate results | • Original photo (no keyword): 2/10  
• Original photo (keywords brain+MRI): 8/10  
• Printed photo (no keyword): 1/10  
• Printed photo (keywords brain+MRI): 7/10  
• LCD screen photo (no keyword): 1/10  
• LCD screen photo (keywords brain+MRI): 6/10 | Average ECG reading score (max. 21):  
• Viewed through GG: 12.4 ± 2.7  
• GG photo on mobile device: 13.4 ± 4.4  
• Original ECG on paper: 14.8 ± 4.4 | (p = 0.45 between GG and mobile device, p = 0.15 between GG and paper, and p = 0.38 between mobile device and paper) |
| Stetler et al, 2015 (Cardiology) | Average ECG reading score (max. 21):  
• Viewed through GG: 12.4 ± 2.7  
• GG photo on mobile device: 13.4 ± 4.4  
• Original ECG on paper: 14.8 ± 4.4 | Average coronary angiogram reading score (max. 17):  
• GG recording viewed on iPad: 14.9 ± 1.1  
• GG recording viewed on desktop computer: 15.2 ± 1.8  
• Original angiogram: 15.9 ± 1.1 | (p = 0.06 between GG recording on iPad and original, p = 0.51 between GG recording on iPad and desktop, and p = 0.43 between GG recording on desktop and original) |
| Duong et al, 2015 (Cardiology) | Average ECG reading score (max. 21):  
• Viewed through GG: 13.5 ± 1.8  
• GG photo on mobile device: 16.1 ± 2.6  
• Original ECG on paper: 18.3 ± 1.7  
• High-resolution camera photo on mobile device: 18.6 ± 1.5 | Average coronary angiogram reading score (max. 17):  
• GG recording viewed on iPad: 14.9 ± 1.1  
• GG recording viewed on desktop computer: 15.2 ± 1.8  
• Original angiogram: 15.9 ± 1.1 | (p = 0.0005 between GG and GG image on mobile device, p = 0.0005 between GG and paper, and p = 0.002 between GG image on mobile device and paper, p = 0.5625 between camera image on mobile device on mobile device and paper, p = 0.125 between camera image on mobile device and GG image on mobile device) |
| Jeroudi et al, 2014 (Cardiology) | Average ECG reading score (max. 21):  
• Viewed through GG: 13.5 ± 1.8  
• GG photo on mobile device: 16.1 ± 2.6  
• Original ECG on paper: 18.3 ± 1.7  
• High-resolution camera photo on mobile device: 18.6 ± 1.5 | Average coronary angiogram reading score (max. 17):  
• GG recording viewed on iPad: 14.9 ± 1.1  
• GG recording viewed on desktop computer: 15.2 ± 1.8  
• Original angiogram: 15.9 ± 1.1 | (p = 0.0005 between GG and GG image on mobile device, p = 0.0005 between GG and paper, and p = 0.002 between GG image on mobile device and paper, p = 0.5625 between camera image on mobile device on mobile device and paper, p = 0.125 between camera image on mobile device and GG image on mobile device) |
| Vallurupalli et al, 2013, (Student training – medical students (cardiology)) | Scenario 1: the trainee wearing GG studied and interpreted the ECG with the senior fellow, subtle signs of ischemia were discussed, and appropriate treatment was initiated.  
Scenario 2: the cardiology fellow wore GG to contact his senior fellow who reviewed the ECG images in question, confirmed that the patient was in cardiac tamponade, and initiated appropriate treatment.  
Scenario 3: the first-year cardiology fellow wore GG to perform an ICD interrogation with the guidance of a senior fellow, a correct diagnosis of ventricular tachycardia was made, and the patient was admitted for appropriate treatment.  
Scenario 4: the cardiology fellow wore GG while performing a subclavian vein cannulation with the attending sitting outside the room and watching the first-person video, allowing the fellow to gain confidence on his own and the attending to assess the fellow. | Average coronary angiogram reading score (max. 17):  
• GG recording viewed on iPad: 14.9 ± 1.1  
• GG recording viewed on desktop computer: 15.2 ± 1.8  
• Original angiogram: 15.9 ± 1.1 | (p = 0.0005 between GG and GG image on mobile device, p = 0.0005 between GG and paper, and p = 0.002 between GG image on mobile device and paper, p = 0.5625 between camera image on mobile device on mobile device and paper, p = 0.125 between camera image on mobile device and GG image on mobile device) |
| Benninger, 2015 (Radiology) | Class success range (ability to wear GG while using SonicEye ultrasound finger probe to identify a structure/space during a 1 minute laboratory examination: | Average ECG reading score (max. 21):  
• Viewed through GG: 13.5 ± 1.8  
• GG photo on mobile device: 16.1 ± 2.6  
• Original ECG on paper: 18.3 ± 1.7  
• High-resolution camera photo on mobile device: 18.6 ± 1.5 | (p = 0.0005 between GG and GG image on mobile device, p = 0.0005 between GG and paper, and p = 0.002 between GG image on mobile device and paper, p = 0.5625 between camera image on mobile device on mobile device and paper, p = 0.125 between camera image on mobile device and GG image on mobile device) |
<table>
<thead>
<tr>
<th>Zahl et al, 2016 (Student training – dental students)</th>
<th>Overall performance:</th>
</tr>
</thead>
<tbody>
<tr>
<td>The initial lab examination score was 81% but improved across six examinations to 97%</td>
<td>There was not a statistically significant difference between mean scores of GG and static camera videos on the VRAES ($t_{22} = 1.702, p = 0.103$)</td>
</tr>
<tr>
<td>Students’ reported mean score on the VRAES was higher for the GG recordings ($\bar{x} = 84.61$) compared to the static camera recordings ($\bar{x} = 79.74$)</td>
<td>Students’ reported mean score on the VRAES was higher for the GG recordings ($\bar{x} = 84.61$) compared to the static camera recordings ($\bar{x} = 79.74$)</td>
</tr>
<tr>
<td>- Students reported that verbal communication was more easily assessed using GG video ($\bar{x} = 23.87$) than static camera video ($\bar{x} = 22.17$) ($p = 0.048$)</td>
<td>- There was a statistically significant difference in mean scores between the GG and static camera videos on the VRAES (p = 0.044), with students reporting that it was easier to assess non-verbal communication using the static camera video ($\bar{x} = 19.78$) than GG ($\bar{x} = 17.09$)</td>
</tr>
<tr>
<td>- Students rated the GG video higher for every item on the verbal communication subscale</td>
<td>- A higher percentage of students rated the static camera video as more effective for self-/peer-assessment of non-verbal communication on every subscale item except ‘eye contact with the patient’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Feng et al, 2015 (Diagnostics – HIV or cancer)</th>
<th>Qualitative decision evaluation using HIV RDTs:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The validation achieved correct diagnosis on all 281 images of OraQuick ADVANCE Rapid HIV-1/2 antibody tests for up to even 5x dilution factors</td>
<td>- The validation achieved correct diagnosis on all 281 images of OraQuick ADVANCE Rapid HIV-1/2 antibody tests for up to even 5x dilution factors</td>
</tr>
<tr>
<td>- A second validation of 145 images was done, again achieving 100% accuracy</td>
<td>- A second validation of 145 images was done, again achieving 100% accuracy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Quantitative decision evaluation using PSA RDTs:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- There was a linear relationship between PSA concentration (ng/mL) and the test line intensity values read for both free and total PSA</td>
<td>- There was a linear relationship between PSA concentration (ng/mL) and the test line intensity values read for both free and total PSA</td>
</tr>
<tr>
<td>- In the free PSA experiments, concentrations above 20 ng/mL start saturating and break the linearity of the response curves</td>
<td>- In the free PSA experiments, concentrations above 20 ng/mL start saturating and break the linearity of the response curves</td>
</tr>
<tr>
<td>- Additionally, for 25 images GG was used to detect and evaluate all PSA tests with minimal differences in test line intensity compared to single RDT images</td>
<td>- Additionally, for 25 images GG was used to detect and evaluate all PSA tests with minimal differences in test line intensity compared to single RDT images</td>
</tr>
<tr>
<td>Reference</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Spencer et al, 2014 (Pulmonology – airway assessment for burn victims)</td>
<td>In both cases, GG was used to document anesthetic airway management in real time, with minimal disruption to clinical care, under standard operating room settings.</td>
</tr>
</tbody>
</table>
| Tully et al, 2015 (Student training – medical students (hospice)) | GG was successfully used to record 30 student/SP encounters:  
  - There was one instance of hardware malfunction, but no other technical issues  
  - Each 9-12-minute session used about 25% of the GG battery life  
  - The built-in microphone did an excellent job picking up the voice of the SPs but was less successful picking up the lower-volume voice of the students |
| Albrecht et al, 2014 (Pathology – autopsy and postmortem examinations) | Average scores of Likert evaluation (1 = low to 5 = high) by 6 raters for images (102 DSLR, 192 GG) taken during autopsy examinations:  
  - Region of interest: 4.45 (DSLR), 3.81 (GG) (p < 0.001)  
  - Sharpness: 4.59 (DSLR), 3.58 (GG) (p < 0.001)  
  - Color: 4.25 (DSLR), 3.80 (GG) (p < 0.001)  
  - Brightness: 3.98 (DSLR), 3.64 (GG) (p = 0.001)  
Average scores ofLikert evaluation (1 = low to 5 = high) by 6 raters for images (102 DSLR, 192 GG) taken during postmortem examinations:  
  - Region of interest: 4.37 (DSLR), 3.77 (GG) (p < 0.001)  
  - Sharpness: 4.25 (DSLR), 4.22 (GG) (p = 0.56)  
  - Color: 4.20 (DSLR), 4.16 (GG) (p = 0.59)  
  - Brightness: 4.13 (DSLR), 4.02 (GG) (p = 0.01) |
| Aldaz et al, 2015 (Chronic Wounds) | Barcode scanning for patient identification:  
  - There was a statistically significant preference for hands-free barcode scanning for routine clinical care (Z(15) = -3.873, p < 0.001, r = 0.71)  
  - All nurses successfully used the SnapCap Glassware to read the patient barcode within four seconds  
Voice-based documentation through video:  
  - Voice-based documentation through a brief video was strongly preferred by nurses (Z(15) = -2.84, p = 0.005, r = 0.52)  
  - Only 12.5% (N = 2/16) of nurses succeeded in using voice commands to launch the video recording feature on their first try  
Double blinking to take photographs:  
  - There was a statistically significant preference for this feature (Z(15) = -3.606, p < 0.001, r = 0.71)  
  - One nurse accidentally took extra photos because GG registered her natural blinking as double blinking  
Head tilt for zooming:  
  - This was the least preferred hands-free interaction method (Z(10) = -1.897, p = 0.058)  
Speech-to-text annotation:  
  - 87.5% (N = 14/16) of nurses successfully saved the two wound
As a result of poor network connectivity, 12.5% (N = 2/16) of nurses required more than ten attempts to successfully save the two wound annotations.

<table>
<thead>
<tr>
<th>Authors, Year, Journal</th>
<th>Details</th>
</tr>
</thead>
</table>
| Chai et al, 2015 (Toxicology) | 18 of 19 attempted GG consults were successfully completed:  
• 89% (N = 16/18) of consults were considered successful by the supervisory consultant  
• Interruptions in the video feed occurred during four consults  
• Interruptions in the audio feed occurred during one consult but did not affect the quality of the consult  
• One consult had interruptions of video and audio feed, thus the consult was unusable  
• The confidence of the supervisory consultants in diagnosing a specific toxidrome in the poisoned patient increased from 59% (N = 10) to 94% (N = 17) after a virtual exam using GG |
| Chai et al, 2014 (Dermatology) | 91% (N = 31/34) attempted connections were completed and considered successful. |

GG: Google Glass; CT: computerized tomography; IQR: interquartile range; LCD: liquid crystal display; MRI: magnetic resonance imaging; ECG: electrocardiogram; ICD: implantable cardiac defibrillator; HIV: human immunodeficiency virus; RDT: rapid diagnostic test; PSA: prostate specific antigen; SP: simulated patient; DSLR: digital single-lens reflex.

3
4
5
6
7
8
9