A comparative study on the psychometric properties of a paper v.s. app administered resilience scale in Scottish youth.

Abstract

Background:
Adequately measuring resilience is important in order to support young people and children who are accessing support through social work or educational settings. A widely accepted measure of youth resilience has been developed by Ungar and Liebenberg which has been shown to work within vulnerable youth [1]. While the measure is completed by the young person on paper, it has been designed to be worked through with a teacher or social worker should further clarification be required. However, this method is time consuming and when faced with large assessment needs can be overwhelming for schools and practitioners. The current study assesses app software with a built-in avatar who can guide the young person through the assessment and its interpretation.

Objective:
The primary objective is to compare the reliability and psychometric properties of a mobile software app to a paper version of the Child and Youth Resilience measure (CYRM-28). Second, the study will assess the use of the CYRM-28 in a Scottish youth population (11-18 years).

Methods:
Following focus groups and discussion with teachers, social workers and young people an avatar was developed by a software company and integrated into an android smartphone app designed to ask questions via the text-to-voice engine built into the device. Seven-hundred and fourteen students from 2 schools in North East Scotland completed either a paper version or app version of the CYRM-28. A cross-sectional design was used and students completed their allocated version twice with a two-week period in between each testing. All participants could request clarification either from a guidance teacher (paper version) or from the in-built software glossary (app version).

Results: Test and retest correlations reported that the app was significantly better performing than the paper version. Paper (r(303)=.76, P<.001, 95%CI [.78, .89]); App (r(413)=.84, P<.001, 95%CI [.77, .85]). Fisher’s r to z transformation found the difference in the correlations to be statistically significant, Z=-2.97, P <.01. Similarly, Cronbach’s alpha in the app condition was very strong (28 items; α=.92) as was the paper version (28 items; α=.87). Fisher’s r to z transformation found the difference in the correlations to be statistically significant, Z=-3.69, P <.01. A Confirmatory Factor Analysis[2] supported the three-factor solution (individual, relational and contextual) and reported a good model fit (χ2 (15, N= 541) = 27.6, P=0.24).

Conclusions:
Alex, an avatar with an integrated voice guide, increased reliability when measuring resilience compared to a paper version with teacher assistance. The CFA reports similar structure using the avatar when compared against the original validation.

**Keywords:** Resilience; CYRM-28; Psychometrics; App-administration.

**Introduction**

**Resilience**

Resilience has previously been conceptualised as an individual difference, for example early research in the field showed that some children, even when exposed to a chaotic family life or early life stressors (e.g. bereavement) had surprisingly healthy behaviours, for example coping ability [3,4,5]. Indeed, a child with high levels of resilience will be able to overcome stressors to achieve a sense of well-being [6]. Furthermore, Panter-Brick and Leckman [7] in a review paper established a pathway between childhood resilience and adult well-being. However, as work on resilience has progressed it has become increasingly recognised that external factors to the child may also have influence on later personal and academic success [3,4,5].

Luthar, Lyman and Crossman [8] categorised subfactors of resilience into three themes labelled as “Attributes of the individual”, “Family influences” and “Wider social environments”. Ungar [9,10] further expanded on these categories to develop a dynamic concept of resilience that puts society at the centre of a child’s ability to develop resilience and coping strategies. Ungar’s ecological model of resilience is culturally sensitive and whilst it does accept that there is individual differences in coping argues that environment surrounding the individual is crucial in providing appropriate resources. For example, while Ungar’s definition and subsequent measurement includes differential aspects of the ability to maintain friendships, it is also measures whether the young person has been provided with the tools to do so. Ungar further suggests that resilience definitions should reflect both ontological and ecological variability and therefore presents the following:

“In the context of exposure to significant adversity, resilience is both the capacity of individuals to navigate their way to the psychological, social, cultural, and physical resources that sustain their well-being, and their capacity individually and collectively to negotiate for these resources to be provided and experienced in culturally meaningful ways (Ungar [11]).”

In Scotland, the setting for the present study, pupils are currently supported via guidance teachers within a framework set by the Government (GIRFEC = Getting It Right For Every Child) and well-being is conceptualized within SHANARRI. SHANARRI has 8 indicators of well-being (Safe, Healthy, Active, Nurtured, Achieving, Respected, Responsible, Included) [12,13]. Guidance teachers lead the pastoral support for pupils of all ages, generally with approximately 200-250 pupils within their care and with whom they will have Personal and Social learning classes each week plus additional support should it be required [14]. It is within this setting that
well-being, resilience and SHANARRI is measured. While there is a positive perception by pupils and parents of the support offered by guidance teachers this is not consistent with large minority of parents arguing that the system does not support their child [15]. The challenge for schools across Scotland is the government lead initiative in which they are expected to assess risks and vulnerability of each child [16] clearly this is easier to do with an app that is able to measure resilience and well-being easily while engaging each pupil. Furthermore, the system is under strain as funding decreases with the education system reduces the number of guidance teachers [17,18].

**Psychometric measurement using apps**

Ungar and Liebenberg [1,2] developed a scale of resilience that reflected this definition of resilience and expressed in three factors (Individual, relational, contextual). Sample questions are “I cooperate with people around me (individual) and “my caregivers watch me closely” (relational). The questionnaire is designed to be used as a verbally administered questionnaire, conducted by a professional within the setting, with responses measured on a likert scale from 1-7. However this is time consuming and difficult to generalise to larger groups of young people. Further studies have extended verbal administration of the questionnaire to a more traditional paper based version in order to widen participation [19]. However, this clearly loses the verbal aspect of the questionnaire which Ungar [11] argues increases participant’s understanding. Therefore, an alternative possibility to personal administration with each child is to use software that allows questions to be read if the participant requires it.

The current study seeks to address the issue of scalability while retaining the verbal aspect and reducing the need for competent reading skills. A further advantage is the benefit of software based data collection which current research indicates reduces chances of incorrect or missed input and therefore increases validation and reliability [20]. Furthermore, there is evidence that internal consistency and concurrent validity are retained when moving to an app based questionnaire [21,22]. Importantly app based scales have consistently been shown to have higher completion rates amongst studies included in a large scale meta-analysis [23]. However, it cannot be assumed that transferring a paper version to an app version can automatically be accepted though there is growing evidence that the transfer to computer based measures does not result in a loss of psychometric properties [24]. This is by no means universal, for example, when transferring paper-and-pen psychometric questionnaires Booth-Kewley et al found that a level of disinhibition crept in to measures regarding such topics as alcohol consumption and risky sexual behaviours [25]. It is therefore, wise to validate the development of a software based app. It is of crucial importance that this is undertaken when the design of the app moves away from the original scale administration, for example in the current study an avatar is used to deliver the questions. Originally data collection online was designed to mimic closely paper questionnaires however recently research has
explored non-human interaction (Bot) with humans and their tendency to disclose, with evidence that self-disclosure increases with the use of non-human interviewee’s [26].

**Present study**
Theories for our love of smartphones has ranged from the use of Bowlby’s attachment theory, addiction behaviours, and their ability to meet our emotional needs [27,28,29]. Indeed, it has been suggested that even larger portable technology, such as laptops, can be seen to be an extension of our identity and selves given we store memories through photographs and access social media on them [30]. For the present study these processes and dynamics are identified as being drivers in the adolescent and relationship with their technological companions which may be seen as an extension of “self” [31,32]. Furthermore, adolescents, have been described as a hard to reach population for research purposes and therefore a smartphone app such as the one tested in the current paper will increase usability [33,34]. It is proposed that the interaction of the above dynamic will encourages honesty in this population and therefore increase the reliability of the questionnaire as has been found in other studies exploring issues of well-being in hard-to-reach populations [35]. Avatar as a researcher is an emerging concept and early studies show increased trust and openness increasing the reliability and confidence in data when discussing sensitive topics [36]. Identification with avatars and robots happens occurs with both humanoid and non-humanoid avatars, for example even computer driven triangle shapes are perceived to have intentionality or taking the perspective of a non-biological computer agent [37,38]. Therefore it is expected that this study will see improved reliability, increased completion rates and similar psychometric properties retained following validity analysis. Additionally, the study aims to validate the use of the CYRM-28 amongst a Scottish population.
Methods

App Development

Feedback on a number of designs of avatar were gathered from 30 professionals, including social workers, educational psychologists and teachers at the 2015 Pathways to Resilience Conference. The outcome of the discussion was to avoid humanoid looking avatars of similar ages to the participants and to opt for one that would be considered gender neutral. Alex has facial elements that move (eyes and mouth) and uses the speech to text engines of the device that is running the app. Alex moves and bounces in response to screen touches. Further focus groups with young people confirmed that Alex was user friendly, approachable and liked by a wide range of ages of both sexes. All participants in the current study were asked to complete a usability questionnaire following the research questions, the results are discussed below.

Design

Recruitment was done through schools who agreed to take part in trials. Information sheets were sent to parents electronically and parents could additionally access a website about the research and agree via online surveys. A cross-sectional design was used aimed at comparing the performances of pen-and-paper to that of an app based CYRM-28 scale [1]. Two schools included all of their pupils as and classes were randomly designated as either app versus paper with age groups represented in each group. All groups were presented with the scale twice, with a two-week retest design. Data collection was completed in PSE (Personal and Social Education) classes and took approximately 10 minutes for the majority of the students following a short explanation regarding the administration of the scale and reminding the students of their ethical rights. A guidance teacher and a member of the data collection team were present during the session. As with the original CYRM-28 participants could request further information and clarification from the researcher regarding questions (paper version) or an inbuilt glossary which can be accessed when the pupil highlights a word or phrase. All research took place during the second term of the academic term (January to March 2017). A third school took part in one app based data collection during the Summer term (July 2017) under the same conditions as described above, further participation was prevented due to exams. This data is included only in the Confirmatory Factor Analysis.
Participants

The participants were 714 students from two North-East Scotland mixed sex schools aged 11-17 (males=354, females =360). Areas in Scotland are divided into five broad groupings of deprivation (1 being most deprived through to 5, least deprived). School 1 (N=403) includes an area of affluence and the majority of pupils fall into bands 4 and 5 (relative high SES according to the Government’s deprivation bands). School two (N=311) is in an urban setting classed as a high deprivation area (all pupils are classed as being in the top two levels of deprivation). The final school draws from a wide range of SES bands. All three schools are comprehensives and therefore mixed ability schools with sixth forms for pupils aged 16-18. The schools used mixed ability groups and each of the schools have approximately similar numbers on the roll.

Materials

The app version ran on Kindle Fires (HD) disconnected from the internet and other software was not able to be accessed. The app presents the question via ALEX an avatar designed by the software company following focus groups with adolescents of a similar age to the participants. Alex is gender neutral and is displayed in diagram 1 below, along with a typical question. As with the paper version the students were required to respond on a 1-7 likert scale (strongly disagree to strongly agree), giving a possible data range of 28-196 with a high score indicating strong resilience. The scale has previously been found to have good reliability scores (Individual = α=.803; Relational = α=.833; Contextual α=.794) and adequate validity following Exploratory and Confirmatory validity [1]. The project received ethical approval from Liverpool Hope Ethics board and students were required to read a short participation information sheet or screen following a short verbal reminder of their right to withdraw from the researcher; parents had previously consented to their children’s participation. Demographic information and data regarding user experience of the app was collected.

Statistical analysis

For demographic descriptive statistics only results from time 1 was included. SPSS 24 was used for descriptive, correlation of time 1 and time 2 test-retest and Cronbach’s alpha. All data were within acceptable skew and met assumptions. Items in the app condition were grouped and calculated to form three factors according to a priori theory developed by Liebenberg and Ungar [2]. The first factor (individual) was composed of 11 items which were further conceptualised as personal skills, peer support and social skills. The second factor of relationship with caregiver included 7 items divided into physical and psychological care. The final factor was labelled as Contextual and had 3 sub-factors (educational, spiritual and cultural).

Twelve cases were removed SPSS prior to a Confirmatory Factor Analysis following identification as multivariate outliers using the Mahalanobis Distance (MD) method. AMOS 24 was used to complete the CFA using a Maximum Likelihood Model.
Results

Usability results

The app based participants were presented with a number of questions which they could decline to complete. Two-hundred and sixty-two of the pupils took part in this second part of the questionnaire. While the majority of the students are happy with the current support offered, a minority were not, with girls more likely to state positive support and more likely to seek help (table 3). Furthermore, a higher number of males and females would not seek help within the current system.

Table 3: Percentage of pupils who answered negatively\(^1\) to the current support questions

<table>
<thead>
<tr>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the guidance system in school provide you with the support you need?</td>
<td>8.5%</td>
</tr>
<tr>
<td>If you need help do you go to your guidance teacher?</td>
<td>22.3%</td>
</tr>
</tbody>
</table>

\(^1\) Combination of those who answered very unlikely or unlikely

Regarding how usable the app was, the majority of the participants rated the app as easy to very easy to use (87.4%) compared (4.4%) to those who rated it hard or very hard. Additionally, users were positive about their experience regarding interaction with Alex. However, participants were moderately negative with the voice that read the instructions with 31% stating that it needed to be changed. They were also encouraged to leave comments regarding improvements; in this field the most common suggestion was to include a game.

Assessment results

Demographic differences for resilience are reported in Table 1, which shows males and females reporting similar scores and suggests there is little difference in resilience across schools. Resilience score decreased with age, with the youngest pupils aged 11 reporting higher levels (M=113.05, SD=11.85) than those aged 16 plus (M=103.50, SD=15.10). A Pearson’s correlation indicated a significant relationship between age and resilience, (r(720)=.81, P=.006, 95%CI [.02, 2.73]).
Table 1. Summary of the scores for each sample including means, SD and confidence intervals.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Mean</th>
<th>SD</th>
<th>Confidence Intervals at 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>S2</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>Paper Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>126</td>
<td>107.85</td>
<td>104.22</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>50</td>
<td>108.06</td>
<td>103.06</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>76</td>
<td>106.53</td>
<td>105.01</td>
</tr>
<tr>
<td>App Total</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234</td>
<td>183</td>
<td>107.45</td>
<td>105.95</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>97</td>
<td>107.65</td>
<td>106.38</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>84</td>
<td>107.57</td>
<td>105.90</td>
</tr>
</tbody>
</table>

\(^a\) S1=School 1 (Deprivation group = 4,5; \(^b\) S2=School 2 (Deprivation group = 1, 2)
These data were collected from two diverse schools and as resilience has been shown to be related to deprivation, it was expected that there would be a difference in the schools. Whilst School 1 reported higher levels of resilience amongst pupils (M=107.24, SD=12.87) than school 2 (M = 105.79, SD = 13.15), this difference was not significant (t(720)=1.38, p=.18). In the paper version scores on CYRM-28 ranged from 63 to 131 (M=106.98, SD=13.51), while the app version of the test scored 56 -135 (M=106.79, SD=13.62). An independent samples t-test was conducted between the two conditions and reported no significant difference (t(720)=-.632, P=.53, 95%CI [-2.55, 1.31]), suggesting that the app version measures resilience as adequately as the paper version. These data for each of the sub-samples are presented in Table 1.

Psychometric properties

Cronbach’s alpha in the app condition was very strong (28 items; α=.92) as was the paper version (28 items; α=.87). Fisher’s r to z transformation found the difference in the correlations to be statistically significant, Z=-3.69, P <.01. Test-retest results were significant in both conditions although the app version was shown to increase reliability: Paper (r(303)=.76, P <.001, 95%CI [.78, .89]); App (r(413)=.84, P <.001, 95%CI [.77, .85]). Fisher’s r to z transformation found the difference in the correlations to be statistically significant, Z=-2.97, P <.01. Reliability statistics suggest that the app administered version of the questionnaire was more reliable than the paper version.

The three-factor structure of the 28-item CYRM-28, based on the model confirmed by Liebenberg and Unger [2], was estimated using a CFA with Time 1 dataset in Amos 24. A maximum-likelihood estimation CFA model was found to be parsimonious, however the chi-square significant indicates that the model did not adequately fit the data, (χ² (df17, N= 541) = 43.78). As large sample sizes can increase the likelihood of significant chi-square other indices of model fit are of particular interest. Table 2 includes a range of fit indices, all of which are within acceptable parameters.

<table>
<thead>
<tr>
<th>X2</th>
<th>Df</th>
<th>P</th>
<th>GFI</th>
<th>CMIN</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original model</td>
<td>43.8</td>
<td>17</td>
<td>&gt;.01</td>
<td>.94</td>
<td>43.78</td>
<td>.98</td>
</tr>
<tr>
<td>Second model</td>
<td>27.59</td>
<td>15</td>
<td>.24</td>
<td>.98</td>
<td>27.59</td>
<td>.99</td>
</tr>
</tbody>
</table>
Modification indices were examined, and several items were found to have significant shared error variance, including: Relational (Physical) and Contextual (Spiritual); Individual (Personal) and Individual (Peer). An exploration of the items that were included in each of these factors for multicollinearity between the items suggested that no item was so redundant with another item that it could be dropped (e1-e2, Tolerance = 1.00, VIF=1.00; e4-e8, Tolerance -1.00, VIF=1.00). As the shared error variance between all of these pairs of items was conceptually consistent with the domain assessed, a final model was re-specified to free these correlated errors, which was found to fit the data moderately well and increased goodness of fit ($\chi^2$ (15, N= 541) = 27.6, $P$ =0.24), further details of fit can be seen in table 2. The final confirmatory factor analytic model of the CYRM - 28 (figure 1) indicates that the items are strongly correlated within factors rather than across factors, this replicates the findings from the original validity study [2]. Diagram 2 shows the error-covariances added to improve the model goodness of fit, each of these were low (r=.12 and r= -.15).

**Discussion**

**Principal Results**
The aim of the study was to establish the adequacy of an app version of a previously validated paper administration of a scale to measure resilience. The app and the paper versions of the scale presented the text of the questions to the participant with Likert scaled responses. The paper condition allowed pupils to ask staff for support whilst in the app version this was built into the device. The results indicate that the app had significantly better reliability in a test-retest situation and significantly higher internal-consistency as measured by Cronbach’s Alpha. Scores across the demographic groups between paper and app were similar, indicating that the app version matches the paper version CYRM-28 when measuring resilience. The results can also support the secondary aim of testing the validity of the measuring in a Scottish population. Finally, the study supports the use of the CYRM-28 in a Scottish youth population (11-18).
Comparison with Prior Work
Ungar had previously identified that resilience was not simply a function of the individual but that influences of the environment were also important [9]. The Confirmatory Factor Analysis reflected this understanding of resilience and as Liebenberg and Ungar had earlier reported the best fitting model was a three-factor solution (individual, family relationships and contextual). Furthermore, the CYRM-28 was designed to be used with the support of an adult professional (teacher or social worker) [2]. However, while this ensures that young people have understood the questions this is not cost effective and therefore is of use only to small groups of children who have been identified as vulnerable. Additionally as discussed in the introduction, increasingly the pastoral system within Scottish schools is under strain. The current study gave evidence that a sizable percentage of children would not seek support from their guidance teachers. The purpose of the present study was to develop a low-cost scalable version of the scale which depends on an avatar to support understanding and encourages openness in adolescents. As discussed by Palmier-Claus the app increased reliability shown by its high internal consistency and additionally participants were more likely to provide similar responses across time periods when using the app version [40]. Again, research had indicated that the use of the avatar in the app would be a positive experience and this has been supported in the present study [41]. The students who completed the supplementary usability questions were generally positive about the avatar. It can be assumed that while app usage was time limited the participants were able to develop a relationship of trust with ALEX and therefore were open in their responses.

Limitations
The current study sought to explore how effective an avatar was in connecting with young people and collecting data about their home-lives and feelings. The outcomes show that the app performed well at this level of data collection and a proof of concept has been met. However, for ethical reasons it was decided to test this on a general population of young people rather than adolescences who have been identified as vulnerable. Furthermore, while it can be argued that resilience is more observable in people who are facing trauma or difficult situations the CYRM-28 has been used in general populations previously [1,19]. Nonetheless, further research that includes vulnerable participants would be of use.

The app is designed to allow the participant as well as the professional to access information about the pupil. The current study has not explored how either of these interact with the reports that are fed back. Research is currently being undertaken to explore how professionals utilize feedback from an app but another question not answered is how the young person themselves react to instant feedback on an aspect of their psychology. Additionally, a discussion around the use of the app within a broader health and social education setting should be developed. The authors strongly suggest that the app fits best in terms of ongoing curriculum around assessing and developing aspects of well-being. Education practitioners and social workers should be involved in a developing good practice of apps such as the
one outlined above. It is suggested that this forms part of a conversation with guidance teachers and young people rather than the end result of an assessment. To that end future research should consider how assessment apps can also include the ability for the participant to communicate with their guidance teachers, this is particularly interesting given the findings in the current study regarding the reluctance of pupils to approach their teachers.

**Conclusions**
The app technology utilized in the current study has shown strong reliability and validity in the measurement of resilience in young adult populations. The current findings demonstrate the efficacy of moving the CYRM-28 ‘gold-standard’ measure of resilience to an online platform. Benefits of avatar-led questioning in relation to young people’s understanding of resilience are evident, however future work should address how technology can be effectively integrated into existing practitioner-led support services within schools.

**Multimedia appendices**
Files have been archived at [https://osf.io/xe9bd/](https://osf.io/xe9bd/)

**Acknowledgements**
APPA-Wellbeing.com who conceptualized and designed the smartphone app.

**Conflicts of Interest**
None declared

**Abbreviations**
CYRM-28: Child and youth resilience measure
GIRFEC: Getting it right for every child
SHANARRI: Safe, Healthy, Active, Nurtured, Achieving, Respected, Responsible, Independent.

**References**
38. (Müller, Oostendorp, Kühn, Brasse, Dijksterhuis & van Baarenm, 2012)