Introduction
Observational coding is a common task in psychological research. In the field of clinical psychology, it is used to systematically record and measure interactions to evaluate client progress and therapist effectiveness. Traditionally, coding is done using paper-based instruments. However, computer-based instruments have the potential to improve this process (MacLin & MacLin, 2005). The present study compared paper-based and computer-based approaches, using a representative coding scale of medium complexity, the Validation and Invalidation Behavior Coding Scale (VIBCS) (Fruzetti, 2001). Validation involves responses that reflect the acceptance and understanding of another’s private experiences (Linehan, 1997).

Hypotheses
The first hypothesis (H1) stated that subjects using a computer-based approach (V-CODE) would exhibit greater reliability when coding psychotherapy sessions than those using a paper-based version of the VIBCS. The second (H2) predicted that ratings in the V-CODE group would be more accurate (i.e. exhibit greater concurrent validity with expert ratings). In addition, it was hypothesized that subjects in the V-CODE group would report greater satisfaction and a better overall experience (H3).

Methods
Participants
Subjects (n = 48) were recruited via the SUNY Oneonta Psychology Department Experiment Management System. Course credit and a Starbucks gift card incentive were offered. Exclusion criteria included any prior observational coding experience.

Measures
Measures included a demographic questionnaire and a post-test questionnaire, which assessed satisfaction as well as confidence in the subject’s performance. The main dependent variable in the questionnaire, which assessed satisfaction as well as confidence in the subject’s performance.

Demographics
Forty-eight students (68.7% female, 4.3% Black/African American, 12.8% Asian/Pacific Islander, 70.2% Caucasian, 12.8% Hispanic) completed the training protocol and coded three mock therapy session videos.

Training Video
A video was developed to provide a brief introduction to the concepts of coding and validation. Examples of the various levels of validation and invalidation were provided. Participants then performed a practice test that was nearly identical to the experimental task.

Mock Therapy Session Videos
A final set of three videos were selected from a pool of six videos developed to represent validation and invalidation in therapy.

Procedure
Subjects were randomly assigned to either a paper-based or computer-based version of the VIBCS coding scale. After an informed consent procedure, subjects completed the demographic questionnaire. Next, they were shown the training video. Afterwards, they coded the three mock therapy session videos (assigned in random order), using either a paper-based or a computer-based approach. Finally, they were asked to fill out a post-test questionnaire, which was followed by a debriefing.

Results

Demographics
Forty-eight students (68.7% female, 4.3% Black/African American, 12.8% Asian/Pacific Islander, 70.2% Caucasian, 12.8% Hispanic) completed the training protocol and coded three mock therapy session videos.

Reliability

<table>
<thead>
<tr>
<th></th>
<th>Paper</th>
<th>V-CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>Average</td>
</tr>
<tr>
<td>Video 1</td>
<td>0.533</td>
<td>0.963</td>
</tr>
<tr>
<td>Video 2</td>
<td>0.164</td>
<td>0.818</td>
</tr>
<tr>
<td>Video 3</td>
<td>0.398</td>
<td>0.938</td>
</tr>
</tbody>
</table>

Inter-rater reliabilities (IRR) were assessed as two-way random, absolute, single-measures Intra-class Correlation Coefficients (ICC) to evaluate the degree of agreement among coders regarding whether the therapists in the mock videos were validating or invalidating the clients. The resulting ICC’s for the first video showed fair degrees of agreement (Chicchetti, 1994). Contrary to H1, there was no significant difference in Turn-by-Turn ICC’s between the Paper and V-CODE groups (p = 0.67). The second hypothesis was also not supported. It was not meaningful to calculate validities, given the low reliability scores. However, because the first video demonstrated a fair degree of reliability, it may be useful to note accuracy scores (V-CODE M = 0.29; Paper M = 0.27). Reported positive attitudes toward research improved significantly from pre to post-test (t(47) = 2.44, p = .018, d = .35). At post-test, approximately 38% of participants reported a high interest in psychological research, compared with 29% at pre-test. Furthermore, over 65% of subjects reported a high level of satisfaction with their participation. There was no significant difference in satisfaction between the V-CODE (M = 3.90) and paper-based groups (M = 3.78).

Discussion
This study provided support for the effectiveness of a short video-based protocol to train undergraduate students to be effective observational coders. This skill can benefit both students and researchers. The reliability scores for video 1 were within the acceptable range for a single coder on a research project. These results are notable, given the possible comprehensiveness and quality of the training. It would be interesting to evaluate possible carry-over effects or subject fatigue on the reliabilities of the subsequent videos. The positive feedback from subjects is also noteworthy. The hypotheses regarding V-CODE were not supported. It is possible that the use of two computers (one for watching the videos, one for coding), in addition to other limitations of V-CODE, may have been factors. It may be valuable to analyze how closely participants’ ratings compare to expert ratings.

Acknowledgements
This project was funded via a Student Research Grant from SUNY Oneonta. We would like to thank our study participants for their time, enthusiasm, and helpful feedback.

References:

