BEA
Educating for Tomorrow's Media

Routledge
Taylor & Francis Group

# Measuring Social and Emotional Content in Children's Television: An Instrument Development Study

**Claire G. Christensen and Carol M. Myford**

*Few researchers have studied the television program characteristics that effectively facilitate social and emotional learning (SEL) in children. To further this line of investigation we created the SEL in Educational Children's Television (SELECT) rating instrument. SELECT ratings indicate whether an educational television episode presents any of 6 SEL skills using any of 5 pedagogical techniques. In this study, 3 raters used the SELECT to rate 80 episodes. Results from multi-facet Rasch analyses illustrated the SELECT's strong content validity, intra- and inter-rater reliability, and sensitivity. Episodes typically presented SEL content implicitly, emphasizing social and decision-making skills most strongly.*

Social and emotional learning (SEL)—developing skills to build healthy relationships, make informed decisions, and manage emotions (Zins & Elias, 2006)—improves academic and behavioral outcomes for children (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). Television (TV) can facilitate SEL; in a meta-analysis of 34 studies, Mares and Woodard (2005) concluded that prosocial TV improves children's social interactions and altruism while decreasing stereotyping and aggression. However, few researchers have identified the TV episode characteristics that most effectively promote SEL—information vital to program improvement. Impeding researchers' ability to isolate and compare characteristics of SEL TV content is the absence of a systematic way to quantify it. To meet that need we developed an instrument to measure the strength of SEL content in educational TV episodes.

***Claire G. Christensen*** *(M.A., University of Illinois at Chicago) studies social-emotional learning in children, with a focus on the effects of prosocial media. She is a graduate student at the University of Illinois at Chicago.*

***Carol M. Myford*** *(Ph.D., University of Chicago) is an associate professor in the Department of Educational Psychology at the University of Illinois at Chicago, where she teaches courses in assessment, program evaluation, and measurement. Her program of research focuses on scoring issues in large-scale performance and portfolio assessments. She has published studies related to rater training, rubric design, quality control monitoring, improving rater performance, and the detection and measurement of rater effects.*


*DOI: 10.1080/08838151.2013.875024*
*Journal of Broadcasting & Electronic Media 58(1), 2014, pp. 21–41*
*ISSN: 0883-8151 print/1550-6878 online*

The Annenberg Public Policy Center (APPC) created the only existing measure relevant to educational TV (Jordan, Schmitt, & Woodard, 2001). This measure assesses the educational quality of ''educational/informational'' (E/I) series: children's TV programs focusing on educational (e.g., SEL-related or academic) content, aired on broadcast networks per a Federal Communications Commission (FCC) requirement (FCC, 1996). The APPC's instrument operationalizes educational quality in terms of five items: overall educational quality, lesson clarity, lesson integration, lesson involvement, and lesson applicability. Studies using this measure found that most E/I series were of moderate to high educational quality (Jordan et al., 2001; Wilson, Kunkel, & Drogos, 2008). However, because the APPC's instrument was designed to evaluate the broad educational quality of E/I series covering diverse lessons, it cannot provide detailed information about SEL content.

A detailed instrument for describing SEL TV content may serve multiple functions. Media researchers could use the instrument to isolate and compare characteristics of episodes' SEL content. Television program creators could use the instrument formatively, to enhance SEL content. Further, parents and educators could use ratings to select high-quality SEL content.

Two episode characteristics—*SEL skill emphasis* and *pedagogical technique inclusion*—may be especially meaningful indicators of SEL content strength. Episodes may have stronger prosocial effects if they emphasize certain SEL skills, such as altruism (Mares & Woodard, 2005). Pedagogical techniques also may affect SEL content strength. For instance, episodes may promote SEL more effectively if characters verbally describe SEL skills (e.g., Elliott & Vasta, 1970) and use skills in situations that resemble real life (e.g., Friedrich-Cofer, Huston-Stein, Kipnis, Susman, & Clewett, 1979). Thus, an appropriate measure of SEL content strength must assess not only *what skills* the episode emphasizes but also *how* it does so.

## Characterizing SEL Content

We created the Social and Emotional Learning in Educational Children's Television (SELECT) rating instrument to assess what SEL skills a TV episode emphasizes, using which pedagogical techniques.

### *SEL Skills.*

The SEL skills on the SELECT fall into three classes, described in the SEL literature: (1) social awareness and interpersonal interaction skills, (2) decision-making skills, and (3) self-awareness and self-management skills (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2005; Durlak et al., 2011; Payton et al., 2000; Zins & Elias, 2006). This three-class framework also undergirds the Illinois State Board of Education's (ISBE) comprehensive Social Emotional Learning Standards for kindergarten through twelfth grade, the only such standards in the United States (see http://www.isbe.net/ils/social_emotional/standards.htm).

Individual SEL skills on the SELECT were drawn from the Illinois Learning Standards for SEL for kindergarten through grade 3. We selected those that Payton and colleagues (2000, p. 6) had previously identified as key SEL skills, and that CASEL included in their definition of SEL skills (CASEL, 2005, pp. 12–13). The SELECT SEL skills are as follows, grouped by class.

- **Social-awareness and interpersonal skills**
  - **Cooperating/helping:** Multiple characters contribute (e.g., offer their opinions or resources) toward a shared goal.
  - **Naming others' emotions:** One character states an emotion that another might feel (e.g., ''You look sad.'').
  - **Resolving conflicts nonviolently:** Characters resolve conflicting interests intentionally and peacefully (e.g., compromise).
- **Decision-making skills:** A character brainstorms, considers options, or assesses potential consequences (e.g., ''How can I fix this? I could use tape or glue.''). Decision-making can be applied in either social or personal situations.
- **Self-awareness and self-management skills**
  - **Naming one's own emotions:** A character identifies his or her emotions (e.g., ''I'm happy,'') or asks viewers to name their emotions (e.g., ''How do you feel?'').
  - **Managing one's own emotions:** A character uses a nonviolent strategy to manage unwanted emotions (e.g., deep breathing or positive self-talk).

Other literatures also emphasize these SEL skills. For example, the social information-processing model asserts that children's social adjustment depends, in part, on their capacity for emotion identification, emotion management, and decision-making (Crick & Dodge, 1994). Likewise, the emotional intelligence literature highlights the importance of emotion awareness and management (Mayer & Salovey, 1997). Further, researchers have found that prosocial behaviors such as cooperating and resolving conflicts peacefully are associated with popularity and friendship (Cillessen & Bellmore, 2002; Hartup, 1996; Newcomb, Bukowski, & Pattee, 1993).

*Pedagogical Techniques.*

To characterize *how* episodes depicted these SEL skills, we drew pedagogical techniques from the APPC's (Jordan et al., 2001) instrument, Fisch's (2000) capacity model, and Elias and Tobias' (1996) skills-training protocol.

*Fisch's (2000) Capacity Model.* Fisch (2000) theorized that children process television in working memory, which has a finite capacity to manage information. Educational TV programs include both narrative and educational content, which simultaneously vie for viewers' limited working memory capacity. To learn, the viewer must devote sufficient working memory capacity to the educational content.

Fisch (2000) posited several episode characteristics that may decrease demands on working-memory capacity, facilitating deeper processing of educational content.

*Elias and Tobias' (1996) Skills-Training Protocol.* Because Fisch's (2000) capacity model is specific to comprehension, we also used a skills-training framework (Elias & Tobias, 1996) to assess episodes' potential to change behaviors. This classroom-based protocol involves clearly describing a new skill and its uses, teaching the skill in concrete steps, guiding rehearsal and feedback, and encouraging independent skill use. Elias and Tobias integrated these techniques into ''Talking with TJ,'' an empirically supported (Dilworth, Mokrue, & Elias, 2002; Rosenblatt & Elias, 2008), video-based social problem-solving intervention for children.

Using these models and the APPC's (Jordan et al., 2001) measure, we identified several pedagogical techniques likely to promote SEL. We classified the techniques as either implicit or explicit.

*Implicit Pedagogical Techniques.* Implicit pedagogical techniques emphasize an SEL skill without explicit mention of that skill (e.g., characters do not name or explain skills). The SELECT includes three implicit pedagogical techniques:

- **Modeling:** Using the skill so that viewers can see or hear it. We drew this technique from the skills-training protocol (Elias & Tobias, 1996).
- **Realistic depiction:** Using the skill such that a child could conceivably replicate it (e.g., without magic or advanced technology). Both the skills-training protocol (Elias & Tobias, 1996) and the APPC (Jordan et al., 2001) endorsed directly applicable educational content.
- **Skill-plot integration:** Incorporating the skill into characters' goal pursuits. Fisch (2000) theorized that integrating educational content into the program's narrative promotes comprehension, because the viewer processes both as an integrated whole. The APPC's (Jordan et al., 2001) rating instrument also assesses skill-plot integration.

*Explicit Pedagogical Techniques.* Explicit pedagogical techniques involve clear discussion of SEL skills (e.g., naming or explaining the skill). Fisch (2000) theorized that explicit educational content is easier to comprehend because it does not require inferences, which tax working memory capacity. The SELECT includes two explicit pedagogical techniques:

- **Naming:** Using a concise, consistent label for an SEL skill (e.g., ''taking deep breaths''). Elias and Tobias (1996) asserted that naming new skills promotes retention.
- **Encouraged verbalization:** Encouraging viewers to talk about SEL skills (e.g., ''Can you say, 'I'm sorry'?''). This is similar to skill rehearsal, which Elias and Tobias (1996) encourage.

## The Present Study

Here we present details regarding the development of a content-analytic rating instrument, the SELECT. Most TV content analyses quantify directly observable characteristics, such as nudity (Baruh, 2009), gender (Emons, Wester, & Scheepers, 2010), or news topics (Kalyango & Onyebadi, 2012). Similarly, the SELECT allows users to quantify observable SEL content characteristics (i.e., individual pedagogical techniques and SEL skills). Unlike most content-analytic coding schemes, however, ratings of these characteristics are combined to create a single score that provides a measure of SEL content strength, much like the APPC's measure of educational quality (Jordan et al., 2001). Thus, in addition to inter-rater reliability, we investigate some less common but important instrument properties to ensure that each observed characteristic contributes to meaningful measurement of SEL content strength.

### Aim 1: Develop a New Instrument, the SELECT, to Measure the SEL Content of E/I TV Episodes

As part of the instrument-development process, we investigated the reliability and validity of SELECT scores for their intended use and interpretation.

### Aim 2: Use the SELECT to Rate a Sample of E/I Episodes and Describe Their SEL Content in Terms of SEL Skills and Pedagogical Techniques

To accomplish this, we explored two research questions:

$RQ_1$: Which SEL skills do episodes emphasize more strongly?
$RQ_2$: Which pedagogical techniques do episodes use more frequently?

## Method

### Sample Selection and Characteristics

*Target Age Range.*

We selected series targeted to viewers ages 10 and younger, because most E/I programs target this age range (Jordan et al., 2001). Further, this age group is more likely to watch television (Scharrer & Comstock, 2003) and to benefit from SEL programming (Mares & Woodard, 2005) than are others.

*Series Screening.*

From an FCC database in which network TV broadcasters describe their E/I series (http://licensing.fcc.gov/KidVid/public/), we selected E/I programs using four criteria to reduce construct-irrelevant variance.

1. A Chicago-market broadcast network (ABC, CBS, NBC, Fox, PAX, or CW) affiliate must have listed the series as E/I programming for 2010 Quarter 3 (Q3, July 1 through September 30) or Quarter 4 (Q4, October 1 through December 31).
2. The series must target children ages 10 and under. Series targeting children older than age 10 (e.g., 7- to 12-year-olds) screened out.
3. The series must air in English.
4. The broadcaster must have aired the series at least three times during 2010 Q3 or Q4.

Our findings could likely generalize to other media markets because most network broadcasters use E/I series provided by their networks (Schmitt, 1999). For instance, NBC broadcasters in most cities air the same NBC-provided E/I series.

The nine network TV broadcasters in the Chicago market listed 147 E/I series for 2010 Q3 and Q4. After we screened out series that exceeded our target age range ($n = 91$) or did not air in English ($n = 1$) and eliminated repeat listings ($n = 35$), the final sample contained 20 series, described in Table 1. Only network broadcasters must air E/I programs (FCC, 1996), so we did not include cable or public TV series.

*Episode Sample.*

We rated a convenience sample of four episodes per series, for a total of 80 episodes. We recorded some episodes when they aired on TV between January and April 2011. We obtained other episodes online or on DVD. When possible we selected from broadcasters' records of episodes that aired during 2010 Q3 and/or Q4. If such records were unavailable, we randomly selected four episodes from the most recent season available.

*Unit of Analysis.*

The unit of analysis was a 30-minute episode, excluding title sequences, commercial breaks, and credit sequences. When series aired two, 10- to 15-minute mini-episodes per half-hour, we rated both mini-episodes together as one. We did this because we assumed that children would watch both mini-episodes back-to-back, and we wished to eliminate construct-irrelevant variance attributable to episode length. For two series, ''Postman Pat'' and ''Noonbory and the Super 7,'' we could not obtain mini-episodes that aired together, so we randomly selected pairs from the available mini-episodes.

**Table 1**
**Series Sample Characteristics**

| Title | Network | Target Age Range (Years) |
|---|---|---|
| 321 Penguins | ION, NBC | 4–8 |
| Adventures from the Book of Virtues | ION | 4–8 |
| Barbar | ION, NBC | 4–8 |
| Boo | ION | 2–5 |
| Busytown Mysteries | CBS | 3–7 |
| Doodlebops Rockin' Road Show | CBS | 3–8 |
| Jane and the Dragon | ION, NBC | 4–8 |
| Magic School Bus | ION | 4–9 |
| Magical DoReMi | CW | 3–7 |
| Marvin the Tap Dancing Horse | ION | 4–8 |
| My Friend Rabbit | ION | 4–8 |
| Mysteries of Alfred Hedgehog | ION | 6–8 |
| Noonbory and the Super 7 | CBS | 3–6 |
| Pearlie | ION | 4–8 |
| Postman Pat | ION | 4–8 |
| Shelldon | ION, NBC | 4–8 |
| Strawberry Shortcake | CBS | 3–6 |
| Turbo Dogs | ION, NBC | 4–8 |
| Willa's Wild Life | ION, NBC | 4–8 |
| Zula Patrol | ION | 4–8 |

## Rating Instrument

The SELECT measures the strength of SEL content in educational TV for children. It appears in Table 2. Episodes receive higher SEL content strength scores if they (a) present one or a few SEL skills using multiple pedagogical techniques, (b) use one or a few pedagogical techniques to present many SEL skills, or (c) present many SEL skills using many pedagogical techniques. In this way, the SELECT is sensitive both to *depth* (e.g., extensive reinforcement of a few SEL skills) and *variety* (e.g., presentation of many SEL skills) of SEL content. Using the SELECT, raters indicate whether the episode employs any of five pedagogical techniques to promote each of six SEL skills. For each of the six SEL skills, the rater assigns a rating of 0 (*no*) or 1 (*yes*) for each of five pedagogical techniques, yielding a matrix of 30 cells, each containing a rating of either 0 or 1. Therefore, an episode's total raw score can range from 0 to 30. This rating procedure provides a more fine-grained description of an episode's SEL content than simple global ratings of SEL skills or pedagogical techniques.

**Table 2**
**Pedagogical Techniques and SEL Skills Included on the SELECT**

| | SEL Skill | | | | | |
|---|---|---|---|---|---|---|
| Pedagogical Technique | Cooperating/ Helping | Naming Others' Emotions | Resolving Conflict Nonviolently | Decision-Making | Naming One's Emotions | Managing One's Emotions |
| Do characters model this skill? | Y/N | Y/N | Y/N | Y/N | Y/N | Y/N |
| Could children use the skill this way? | Y/N | Y/N | Y/N | Y/N | Y/N | Y/N |
| Is this skill important to the plot? | Y/N | Y/N | Y/N | Y/N | Y/N | Y/N |
| Do characters use the same name for the skill at least twice? | Y/N | Y/N | Y/N | Y/N | Y/N | Y/N |
| Do characters encourage viewers to verbalize information related to this skill? | Y/N | Y/N | Y/N | Y/N | Y/N | Y/N |

*Note.* The SELECT yields a total score (the total number of ''Yes'' responses), along with scores for each SEL skill (the number of ''Yes'' responses in each skill's column) and each pedagogical technique (the number of ''Yes'' responses in each technique's row). On the rater version we arranged the SEL skills and pedagogical techniques in a random sequence to prevent order effects.

In addition to providing a measure of overall SEL content strength for each episode, the SELECT allows researchers to assess how frequently episodes include each pedagogical technique and how strongly episodes emphasize each SEL skill. The note to Table 2 explains how episodes are scored.

The six SEL skills on the SELECT are (a) cooperating/helping, (b) naming others' emotions, (c) resolving conflicts nonviolently, (d) decision-making, (e) naming one's emotions, and (f) managing one's emotions. For some analyses, we grouped the skills into classes: social-awareness and interpersonal skills (a, b, and c), decision-making skills (d), and self-awareness and self-management skills (e and f).

Using the SELECT, raters also indicate which pedagogical techniques an episode employs to emphasize each SEL skill. The SELECT's five pedagogical techniques are (a) modeling, (b) realistic depiction, (c) skill-plot integration, (d) skill naming, and (e) encouraged verbalization. For some analyses we classified the pedagogical techniques as either implicit (a, b, and c) or explicit (d and e).

## Procedure

*Raters.*

Raters were the first author and two undergraduate research assistants (RAs). We intended to refine the instrument rather than test hypotheses; thus, the lead researcher participated in episode rating to facilitate informed revisions of the SELECT. RAs received course credit for participation. All raters were in their 20s; two were female.

*Judging Plan.*

We used a fully crossed judging plan: all three raters rated all 80 episodes.[1] This allowed us to explore how different raters perceived the same SEL skills and pedagogical techniques within various episodes and to refine the instrument accordingly.

*Rater Training and Rating Procedure.*

The lead researcher trained the RAs to use the SELECT. RAs studied the coding manual over 2 weeks and used the SELECT to rate two practice episodes, excluded from the dataset. We compared their practice ratings with the lead researcher's, discussing and resolving discrepancies until over 80% of the ratings were in exact agreement. After training, raters independently rated eight to ten episodes per week from January through April 2011. They assigned ratings while watching each episode, pausing and rewinding as necessary to clarify their understanding. For training and instrument refinement purposes, raters also noted in what scene(s) they noticed each SEL skill.

Each week we analyzed all of the ratings assigned to date. We used the output from this analysis to identify specific ratings (i.e., one rater's ''Yes'' or ''No'' rating to indicate whether an episode used a particular pedagogical technique to emphasize a particular SEL skill) that were inconsistent with the overall rating pattern. Then all three raters met to discuss these inconsistent ratings, changing those that we deemed to be inaccurate. We corrected 74 ratings out of 7,170 total, or 1% of our ratings. This helped us to refine the coding manual and ensure the accuracy of our ratings. This process of resolving rater discrepancies is recommended in qualitative content analyses (Bernard & Ryan, 2010; Forman & Damschroder, 2007); employed in content analyses in psychology and media (e.g., Bender, Thompson, McManus, Lantry, & Flynn, 2007; Griggs, Jackson, Christopher, & Marek, 1999; Jansen & Resnick, 2006); and routinely used by researchers in other fields such as medicine (e.g., Gagliardi & Dobrow, 2011; Harris, Needer, Ellerb, & Bowe, 2011; Hysong et al., 2010).

*Institutional Review Board.*

The Institutional Review Board at the University of Illinois at Chicago confirmed that this project did not include human subjects.

## Multi-Facet Rasch Measurement Model

A strength of this study was our use of multi-facet Rasch measurement (MFRM) analysis, which helped us to refine the SELECT, monitor rater performance, and describe episodes' SEL content. Using MRFM analysis, the researcher can pinpoint unexpected ratings associated with specific raters, pedagogical techniques, and SEL skills (i.e., those ratings that do not ''fit'' with the other ratings that raters assigned to a given SEL skill or pedagogical technique and thus may be inaccurate). The researcher can use this information to locate potential sources of error and refine the rating instrument (and raters' ratings) accordingly. Further, if data demonstrate sufficient fit to the Rasch model, we can directly compare the strength of SEL content in each episode, the leniency of each rater, the strength of emphasis on each SEL skill, and the frequency of use of each pedagogical technique. This facilitates both the monitoring of raters' performance and the description of episodes' SEL content.

We used the Facets (v3.67.0) software to run MRFM analyses on the data (Linacre, 2010). For most analyses, we examined four facets of the data: episode SEL content strength, rater leniency (i.e., a rater's tendency to assign high or low ratings), SEL skill emphasis, and pedagogical technique frequency. We conducted a multi-facet partial credit Rasch analysis, modeling each rating scale for each pedagogical technique separately. For some analyses, we compared classes of SEL skills or pedagogical techniques, rather than individual skills or techniques. See the Appendix for all measurement models.

## Interpreting Rasch-Generated Statistical Indicators

*Rasch Measures.*

Output from a MRFM analysis provides measures (with standard error estimates) of the strength of SEL content for each episode, the leniency of each rater, the strength of emphasis on each SEL skill (and SEL skill class), and the frequency of usage of each pedagogical technique (and pedagogical technique class). For ease of interpretation, we linearly rescaled the Rasch measures so that the resulting scale would run from 0 to 10. All facets are positively oriented. In other words, the closer an episode's SEL content measure is to 10, the more ratings of 1 (i.e., ''Yes'') raters assigned to individual SEL skills and pedagogical techniques for that episode. The same is true for raters, SEL skills, and pedagogical techniques. A higher measure indicates more ratings of 1 (i.e., ''Yes'') for that facet.

*Rasch Fit Indices.*

The *infit mean-square statistic* indexes the degree to which each rater, SEL skill, and pedagogical technique contributes meaningful information to the measurement of SEL content strength. It has an expected value of 1 and can range from 0 to infinity. For rating situations in which agreement is encouraged, as in our study, an acceptable range for this statistic is 0.4 to 1.2 (Wright & Linacre, 1994).

Infit mean-square statistics below 0.4 indicate that ratings are not providing unique, independent data. For example, an infit mean-square statistic less than 0.4 for a rater would indicate that the rater did not rate each pedagogical technique (or SEL skill) independently, but instead assigned many of the same ratings across pedagogical techniques and/or SEL skills, perhaps not being able to distinguish among them. Similarly, an infit mean-square statistic less than 0.4 for an SEL skill or pedagogical technique would indicate that the ratings for that particular skill or pedagogical technique too closely mirrored the ratings of other skills or pedagogical techniques, and thus did not provide independent data that was useful for distinguishing among the episodes in terms of the SEL content strength. That is, the ratings of that particular skill or pedagogical technique were overly redundant with ratings of other skills or pedagogical techniques.

By contrast, infit mean-square statistics above 1.2 indicate that ratings are too inconsistent with the overall rating pattern to contribute meaningfully to the measurement of SEL content strength. For instance, an infit mean-square statistic above 1.2 for a rater would indicate that one or more of the rater's ratings were quite inconsistent with ratings that other raters assigned (Linacre, 2002). Similarly, an infit mean-square statistic above 1.2 for an SEL skill or pedagogical technique would indicate that the ratings for that particular skill or pedagogical technique were not consistent with the ratings of the other skills or pedagogical techniques. Thus, those particular ratings did not 'fit' with the others and consequently were not contributing meaningfully to the measurement of SEL content strength.

We also examined *point-biserial correlation coefficients* for the SEL skills and pedagogical techniques. The point-biserial correlation coefficient indicates the extent to which high ratings for a particular individual SEL skill or pedagogical technique are associated with high total scores for episodes. For example, a positive point-biserial correlation coefficient for the SEL skill of decision-making would indicate that ratings of decision-making were positively correlated with the episodes' total scores; by contrast, a negative point-biserial correlation coefficient would indicate that the ratings of decision-making were negatively correlated with the episodes' total scores.

*Rasch Separation Statistics.*

We examined the separation statistics from our MFRM analyses to determine how reliably SELECT ratings could distinguish episodes with strong SEL content from those with weak content. To be useful, the SELECT would need to credibly order the episodes by their SEL content strength measures. Output from our analyses included an *episode strata estimate*, which indicates the number of statistically distinct strata of episodes within the sample we rated. Additionally, we report an *episode separation reliability*, which indicates how reliable the rank ordering of the episodes by their SEL content strength was. Separation reliability values range from 0 to 1, with values closer to 1 indicating higher reliability.

# Results

## Aim 1: Investigation of the Reliability and Validity of SELECT Scores

To study whether the SELECT was psychometrically sound, we posed four reliability- and validity-related questions.

*Content Validity.*

Does each SEL skill and pedagogical technique included on the SELECT contribute meaningfully to the measurement of SEL content strength? The infit mean-square statistics for the six SEL skills ranged from 0.82 to 1.16 and from 0.95 to 1.08 for the five pedagogical techniques (see Table 3), all within the acceptable range of 0.4 to 1.2 (Wright & Linacre, 1994). In addition, the SEL skills and pedagogical techniques all had positive point-biserial correlation coefficients. These findings suggest that all the SEL skills and pedagogical techniques were working together to measure SEL content strength, providing strong evidence of content validity.

**Table 3**
**Measures, Infit Mean-Square Statistics, Point-Biserial Correlations, and Average Raw Scores for Individual Elements of Three SELECT Facets**

| Individual Elements of Each SELECT Facet | Measure[a] | Infit Mean-Square | Point-Biserial Correlation | Average Raw Score |
|---|---|---|---|---|
| | | SEL skill | | |
| Cooperating/helping | 9.29 (0.15) | 1.01 | 0.42 | 0.56 (0.02) |
| Naming others' emotions | 6.46 (0.13) | 0.82 | 0.38 | 0.34 (0.03) |
| Decision-making | 6.38 (0.13) | 1.16 | 0.32 | 0.33 (0.03) |
| Naming one's emotions | 5.98 (0.13) | 0.88 | 0.35 | 0.30 (0.03) |
| Resolving conflicts | 5.97 (0.13) | 1.00 | 0.33 | 0.30 (0.03) |
| Managing one's emotions | 3.41 (0.16) | 1.08 | 0.18 | 0.12 (0.02) |
| | | SEL skill class | | |
| Social-awareness/ interpersonal | 7.23 (0.07) | 0.95 | 0.37 | 0.40 (0.03) |
| Decision-making skills | 6.50 (0.12) | 1.10 | 0.32 | 0.33 (0.03) |
| Self-awareness/ management | 5.01 (0.09) | 1.00 | 0.27 | 0.21 (0.03) |
| | | Pedagogical technique | | |
| Skill modeling | 9.99 (0.10) | 0.95 | 0.29 | 0.61 (0.05) |
| Skill-plot integration | 9.56 (0.10) | 0.95 | 0.29 | 0.56 (0.06) |
| Realistic skill portrayal | 8.15 (0.10) | 1.03 | 0.26 | 0.40 (0.05) |
| Skill naming | 2.65 (0.23) | 1.08 | 0.08 | 0.04 (0.02) |
| Encouraged verbalization | 0.90 (0.36) | 1.08 | 0.05 | 0.01 (0.01) |
| | | Pedagogical technique class | | |
| Implicit | 9.82 (0.06) | 0.98 | 0.27 | 0.52 (0.05) |
| Explicit | 2.68 (0.19) | 1.07 | 0.07 | 0.03 (0.01) |
| | | Rater | | |
| 3 | 7.04 (0.09) | 0.90 | 0.38 | 0.38 (0.03) |
| 1 | 6.27 (0.09) | 1.07 | 0.34 | 0.33 (0.03) |
| 2 | 5.43 (0.10) | 0.99 | 0.34 | 0.27 (0.03) |

*Note*. Standard errors are in parentheses. SELECT = Social and Emotional Learning in Educational Children's Television rating instrument. SEL = Social and Emotional Learning.
[a]Measures for SEL skills and SEL skill classes represent SEL skill emphasis. Measures for pedagogical techniques and pedagogical technique classes represent pedagogical technique frequency. Measures for raters represent rater leniency.

*Intra-Rater Reliability*.

Can each rater rate the SELECT skills and pedagogical techniques in an internally consistent manner? The infit mean-square statistics were 1.07 for Rater 1, 0.99 for Rater 2, and 1.12 for Rater 3, all within the acceptable range of 0.4 to 1.2 (Wright & Linacre, 1994). Thus each rater was able to use the SELECT in an internally consistent fashion.

*Inter-Rater Reliability*.

Can multiple raters use the SELECT in a similar fashion? Despite small differences in rater leniency (see Table 3), the three raters demonstrated 81.8% exact agreement in their ratings, exceeding the model expectation of 75.5%. Additionally, Fleiss' kappa was .59. These findings indicate that the raters could use the SELECT in a similar manner, assigning ratings that exhibited a satisfactory level of inter-rater reliability. Furthermore, the fact that the infit mean-square statistics for the individual SEL skills and the pedagogical techniques were all within the acceptable range of 0.4 to 1.2 indicates that the raters were able to code each of the SEL skills and pedagogical techniques in a reliable manner, after correcting for chance agreement.

*Sensitivity*.

Can SELECT scores reliably distinguish episodes with strong SEL content from those with weak SEL content? The episode strata estimate was 3.43; that is, there were about three-and-a-half statistically distinct levels of SEL content strength among the 80 episodes we rated. The episode separation reliability was .84, indicating that the SELECT produced a reliable rank ordering of those episodes by their SEL content strength scores, consistently distinguishing strong from weak SEL content. These findings provide strong evidence of the SELECT's sensitivity and its practical utility for rank ordering episodes by their SEL content strength scores.

## Aim 2: SEL Content in Educational/Informational Episodes

Because our analyses indicated that SELECT scores are valid and reliable, we then addressed Aim 2, using SELECT scores to describe episodes' SEL content in terms of SEL skills and pedagogical techniques.

*$RQ_1$: Which SEL Skills do Episodes Emphasize More Strongly?*

The episodes tended to emphasize some SEL skill classes more than others, as indicated by results from the fixed chi-square test, $\chi^2(2) = 353.7^2$, $p < .01$. The measures for all three SEL skill classes were significantly different from one another:[3] self-awareness/management and decision-making skills, $t(79) = 9.93$, $p < .001$; decision-making and social-awareness/interpersonal skills, $t(79) = 5.25$, $p < .001$; and self-awareness/management and social-awareness/interpersonal skills, $t(79) =$

19.47, $p < .001$. The episodes emphasized social-awareness/interpersonal skills strongly, decision-making skills moderately, and self-awareness/management skills weakly (see Table 3).

Next, we examined measures for individual SEL skills (shown in Table 3) to identify which specific skills episodes emphasized most strongly. Again, the fixed chi-square test for the SEL skills was significant, $\chi^2(5) = 770.9$, $p < .01$, indicating that episodes tended to emphasize some SEL skills more than others. Among the social-awareness/interpersonal skills, episodes emphasized cooperating/helping more strongly than naming others' emotions, $t(79) = 14.26$, $p < .001$; and naming others' emotions more strongly than nonviolent conflict resolution, $t(79) = 2.67$, $p < .01$. Among the self-awareness/management skills, episodes emphasized naming one's emotions more strongly than managing one's emotions, $t(79) = 12.47$, $p < .001$.

*$RQ_2$: Which Pedagogical Techniques do Episodes use More Frequently?*

Episodes were more likely to use implicit pedagogical techniques than explicit pedagogical techniques, as indicated by results from the fixed chi-square test, $\chi^2(1) = 1319.1$, $p < .001$ (see Table 3). Among the implicit pedagogical techniques, episodes used skill modeling more frequently than skill-plot integration, $t(79) = 3.04$, $p < .01$; and episodes used skill-plot integration more frequently than realistic skill portrayal, $t(79) = 9.97$, $p < .001$. Between the explicit pedagogical techniques, episodes used skill naming more than encouraged verbalization, $t(79) = 4.10$, $p < .001$.

## Discussion

Few researchers have explored the mechanisms by which prosocial TV programs enhance children's SEL. To facilitate this work we created the SELECT rating instrument to yield detailed information about which SEL skills episodes emphasize using which pedagogical techniques. The SELECT is psychometrically sound; we found strong evidence of content validity, intra- and inter-rater reliability, and instrument sensitivity. This instrument facilitates a more detailed analysis of SEL TV content than any existing measure, with applications for research, program improvement, and quality control.

### Applications

*Furthering Research.*

SELECT ratings only describe episodes' SEL content; they do not measure the extent to which a TV episode will promote viewers' SEL. However, rigorous quan-

tification of SEL content, using an instrument like the SELECT, is the first step in determining what characteristics of SEL TV content are most educational, and for whom. For example, we found that episodes often emphasized naming others' emotions and making well-reasoned decisions, yet little is known about whether TV can effectively promote these skills. Researchers could use the SELECT to identify TV episodes that emphasize these skills, and then investigate whether these episodes enhance viewers' emotion-naming and decision-making abilities.

In addition, researchers could use the SELECT to determine whether specific types of SEL content are more likely to promote SEL skills among certain children. For instance, Tower, Singer, Singer, and Biggs (1979) found that children who had higher IQs showed greater gains in cooperation and interaction with adults after watching *Mister Rogers' Neighborhood* than did children who had lower IQs. By contrast, children who had lower IQs showed greater gains in those behaviors after watching *Sesame Street* than did children who had higher IQs. The SELECT could help researchers identify differences in the SEL content of these two programs that might explain these participant-characteristic interactions. Research using the SELECT could be helpful in identifying the types of SEL content that most effectively teach children at various stages of development.

*Promoting Program Improvement.*

Television program creators could use the SELECT formatively, to assess strengths and weaknesses of episodes' SEL content and make improvements. For example, our analyses highlighted some common strengths and weaknesses in the pedagogical techniques episodes used to promote SEL. Strengths included frequent use of modeling—thought to promote skill learning (Elias & Tobias, 1996), and skill-plot integration—theorized to enhance comprehension (Fisch, 2000). Meanwhile, according to Fisch's (2000) capacity model, episodes' infrequent use of explicit pedagogical techniques presents a significant weakness, which program creators could remediate by including more SEL-skill naming in episodes. Doing so may promote retention (Elias & Tobias, 1996).

*Facilitating Quality Control.*

The SELECT could also facilitate quality control, allowing policymakers, educators, and parents to discern which E/I programs include strong SEL content. Most E/I programs tend to focus on SEL content, rather than other educational subjects (Jordan et al., 2001). Our results suggest that the strength of this SEL content varies by episode—we detected over three statistically distinct levels of SEL content strength in this sample. Thus the SELECT may be one useful tool for assessing how ''educational/informational'' E/I episodes truly are. In addition, parents and educators might use the SELECT to choose strong or relevant SEL TV content. For

instance, educators might use SELECT ratings to identify a TV episode that will strongly reinforce a particular SEL standard.

*Using the SELECT.*

We emphasize the importance of training raters to apply the coding scheme and subsequently monitoring rater performance. Before they begin rating operationally, new raters should study the coding manual (Christensen, 2013), rate sample episodes, receive feedback on their ratings, and discuss discrepancies to rectify misunderstandings. Once rating operationally, raters should meet periodically to discuss their ratings as a means of maintaining accuracy.

## Limitations

Our primary aim was to develop a rating instrument and illustrate its use, not to provide a generalizable description of all E/I content. Thus the results we obtained from our convenience sample may not generalize to all E/I episodes. The episodes we rated may have been available on Web sites and DVDs because producers felt they were exemplary. However, Woodard (1999) previously found no difference in the educational quality of E/I episodes researchers taped off the air versus episodes that broadcasters selected for review. In addition, for two series we created artificial episodes, combining randomly selected mini-episodes (see Method section), which may not be representative of typical episodes for these series. Additional research is needed to assess whether our findings generalize to E/I programs for older children and programs airing on cable or public television.

The SELECT includes limited SEL skills and pedagogical techniques. However, the literature supports both the SEL skills (CASEL, 2005; Durlak et al., 2011; Payton et al., 2000; Zins & Elias, 2006) and pedagogical techniques (Elias & Tobias, 1996; Fisch, 2000; Jordan et al., 2001) included in the SELECT.

Fleiss' kappa for this study was moderate. However, exact agreement was high. Exact agreement may be a more valid measure of interrater reliability, as Fleiss' kappa tends to fluctuate with sample base rates (Uebersax, 1987).

# Conclusions

The SELECT is the first rating instrument designed to measure TV episodes' SEL content in detail. Media researchers, program creators, and educators might find this instrument useful for studying, improving, and choosing strong SEL TV content.

# Appendix

## Measurement Models

### Measurement Model for the Many-Facet Rasch Analysis of Individual SEL Skills and Pedagogical Techniques

$$\log [P_{nidjk}/P_{nidj(k-1)}] = B_n - D_i - S_d - C_j - F_{ik}$$

where

$P_{nidjk}$ = the probability that episode $n$ will receive a rating of $k$ from rater $j$ on pedagogical technique $i$ for SEL skill $d$,
$P_{nidj(k-1)}$ = the probability that episode $n$ will receive a rating of $k - 1$ from rater $j$ on pedagogical technique $i$ for SEL skill $d$,
$B_n$ = the strength of SEL content in episode $n$,
$D_i$ = the frequency of use of pedagogical technique $i$,
$S_d$ = the strength of emphasis on SEL skill $d$,
$C_j$ = the leniency of rater $j$, and
$F_{ik}$ = the difficulty of scale category $k$, relative to scale category $k - 1$ for pedagogical technique $i$.

### Measurement Model for the Many-Facet Rasch Analysis of SEL Skill Classes

$$\log [P_{nidjk}/P_{nidj(k-1)}] = B_n - D_i - K_d - C_j - F_{ik}$$

$P_{nidjk}$ = the probability that episode $n$ will receive a rating of $k$ from rater $j$ on pedagogical technique $i$ for a SEL skill included in skill class $d$,
$P_{nidj(k-1)}$ = the probability that episode $n$ will receive a rating of $k - 1$ from rater $j$ on pedagogical technique $i$ for a SEL skill included in skill class $d$,
$B_n$ = the strength of SEL content in episode $n$,
$D_i$ = the frequency of use of pedagogical technique $i$,
$K_d$ = the strength of emphasis on skills in SEL skill class $d$,
$C_j$ = the leniency of rater $j$, and
$F_{ik}$ = the difficulty of scale category $k$, relative to scale category $k - 1$ for pedagogical technique $i$.

### Measurement Model for the Many-Facet Rasch Analysis of Pedagogical Technique Classes

$$\log [P_{nidjk}/P_{nidj(k-1)}] = B_n - L_i - S_d - C_j - F_{ik}$$

$P_{nidjk}$ = the probability that episode $n$ will receive a rating of $k$ from rater $j$ on a pedagogical technique included in technique class $i$ for SEL skill $d$,
$P_{nidj(k-1)}$ = the probability that episode $n$ will receive a rating of $k - 1$ from rater $j$ on a pedagogical technique included in technique class $i$ for SEL skill $d$,
$B_n$ = the strength of SEL content in episode $n$,
$L_i$ = the frequency of use of pedagogical techniques in technique class $i$,
$S_d$ = the strength of emphasis on SEL skill $d$,
$C_j$ = the leniency of rater $j$, and
$F_{ik}$ = the difficulty of scale category $k$, relative to scale category $k - 1$ for pedagogical technique class $i$.

## Notes

[1]We lost Rater 2's ratings for one episode due to a failure of our data management system.

[2]We performed all the chi-square tests and *t*-tests on the measures obtained from MFRM analyses. For instance, to determine which SEL skill class the episodes emphasized most frequently, we compared the three SEL skill emphasis measures for social, decision-making, and personal SEL skills, as reported in Table 3. In this case, the chi-square test is testing the hypothesis that the three SEL skill emphasis measures are not statistically significantly different from one another.

[3]To minimize the risk of Type 1 errors, we report Holm-Bonferonni-corrected $p$ values (Holm, 1979) throughout.

## References

Baruh, L. (2009). Publicized intimacies on reality television: An analysis of voyeuristic content and its contribution to the appeal of reality programming. *Journal of Broadcasting & Electronic Media*, *53*, 190–210.

Bender, K., Thompson, S. J., McManus, H., Lantry, J., & Flynn, P. M. (2007). Capacity for survival: Exploring strengths of homeless street youth. *Child Youth Care Forum*, *36*, 25–42.

Bernard, H. R. & Ryan, G. W. (2010). *Analyzing qualitative data: Systematic approaches*. Thousand Oaks, CA: Sage.

Christensen, C. G. (2013). *SELECT coding manual*. Unpublished Manuscript, Department of Psychology, University of Illinois at Chicago, Chicago, IL.

Cillessen, A. H., & Bellmore, A. D. (2002). Social skills and interpersonal perception in early and middle childhood. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 355–374). Malden, MA: Blackwell.

Collaborative for Academic, Social, and Emotional Learning. (2005). *Safe and sound: An educational leader's guide to evidence-based social and emotional learning programs—Illinois edition*. Chicago, IL: Author.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74.

Dilworth, J. E., Mokrue, K., & Elias, M. J. (2002). The efficacy of a video-based teamwork-building series with urban elementary school students: A pilot investigation. *Journal of School Psychology, 40,* 329–346.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432.

Elias M. J., & Tobias, S. E. (1996). *Social problem-solving interventions in the schools*. New York, NY: Guilford.

Elliott, R., & Vasta, R. (1970). The modeling of sharing: Effects associated with vicarious reinforcement, symbolization, age, and generalization. *Journal of Experimental Child Psychology, 10,* 8–15.

Emons, P., Wester, F., & Scheepers, P. (2010). ''He works outside the home; She drinks coffee and does the dishes:'' Gender roles in fiction programs on Dutch television. *Journal of Broadcasting & Electronic Media*, *54*, 40–53.

Federal Communications Commission. (1996). *Policies and rules concerning children's television programming: Revision of programming policies for television broadcast stations* (MMDocket No. 93-48).

Fisch, S. M. (2000). A capacity model of children's comprehension of educational content on television. *Media Psychology, 2*, 63–91.

Forman, J., & Damschroder, L. (2007). Qualitative content analysis. In L. Jacoby and L. A. Siminoff (Eds.), *Empirical methods for bioethics: A primer* (Advances in Bioethics, Vol. 11, pp. 39–62). Bingley, United Kingdom: Emerald.

Friedrich-Cofer, L. K., Huston-Stein, A., Kipnis, D. M., Susman, E. J., & Clewett, A. S. (1979). Environmental enhancement of prosocial television content: Effects on interpersonal behavior, imaginative play, and self-regulation in a natural setting. *Developmental Psychology, 16,* 637–646.

Gagliardi, A. R., & Dobrow, M. J. (2011). Paucity of qualitative research in general medical and health services and policy research journals: Analysis of publication rates. *BMC Health Services Research*, *11*, 268–275.

Griggs, R. A., Jackson, S. L., Christopher, A. N., & Marek, P. (1999). Introductory psychology textbooks: An objective analysis and update. *Teaching of Psychology*, *26*, 182–189.

Harris, A. H. S., Needer, R. N., Ellerb, L. S., & Bowe, T. R. (2011). Validation of the treatment identification strategy of the HEDIS addiction quality measures: Concordance with medical record review. *BMC Health Services Research, 11*, 73.

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67*, 1–13.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6,* 65–70.

Hysong, S. J., Sawhney, M. K., Wilson, L., Sittig, D. F., Espadas, D., Davis, T., & Singh, H. (2010). Provider management strategies of abnormal test result alerts: A cognitive task analysis. *Journal of the American Medical Informatics Association, 17,* 71–77.

Illinois State Board of Education. (n.d.). *Illinois learning standards: Social/emotional learning* (*SEL*). Springfield, IL: Author. Retrieved from http://www.isbe.net/ils/social_emotional/standards.htm

Jansen, B., & Resnick, M. (2006). An examination of searcher's perceptions of nonsponsored and sponsored links during ecommerce Web searching. *Journal of the American Society for Information Science and Technology, 57*, 1949–1961.

Jordan, A. B., Schmitt, K. L., & Woodard, E. H. (2001). Developmental implications of commercial broadcasters' educational offerings. *Applied Developmental Psychology, 22*, 87–101.

Kalyango, Y., & Onyebadi, U. (2012). Thirty years of broadcasting Africa on U.S. network television news. *Journal of Broadcasting & Electronic Media*, *56*, 669–687.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*, 878.

Linacre, J. M. (2010). *FACETS* (*Version 3.67.0*) [Computer software]. Chicago, IL: Winsteps® Rasch Measurement Software.

Mares, M. L., & Woodard, E. (2005). Positive effects of television on children's social interactions: A meta-analysis. *Media Psychology, 7*, 301–322.

Mayer J. D., & Salovey P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence* (pp. 3–31). New York, NY: Basic Books.

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin, 113*, 99–128.

Payton, J. W., Wardlaw, D. M., Graczyk, P. A., Bloodworth, M. R., Tompsett, C. J., & Weissberg, R. P. (2000). Social and emotional learning: A framework for promoting mental health and reducing risk behaviors in children and youth. *Journal of School Health, 70*, 179–185.

Rosenblatt, J. L., & Elias, M. J. (2008). Dosage effects of a preventive social-emotional learning intervention on achievement loss associated with middle school transition. *Journal of Primary Prevention, 29,* 535–555.

Scharrer, E., & Comstock, G. (2003). Entertainment televisual media: Content patterns and themes. In E. L. Palmer & B. M. Young (Eds.), *The faces of televisual media: Teaching, violence, selling to children* (pp. 161–193). Mahwah, NJ: Erlbaum.

Schmitt, K. (1999). *The three-hour rule: Is it living up to expectations?* (Report No. 30). Philadelphia, PA: The Annenberg Public Policy Center of the University of Pennsylvania.

Tower, R. B., Singer, D. G., Singer, J. L., & Biggs, A. (1979). Differential effects of television programming on preschoolers' cognition, imagination, and social play. *American Journal of Orthopsychiatry, 49,* 265–281.

Uebersax, J. S. (1987). Diversity of decision-making models and the measurement of interrater agreement. *Psychological Bulletin*, *101*, 140–146.

Wilson, B., Kunkel, D., & Drogos, K. L. (2008). *Educationally/Insufficient? An analysis of the availability and educational quality of children's E/I programming*. Oakland, CA: Children Now.

Woodard, E. (1999). *The 1999 state of children's television report: Programming for children over broadcast and cable television* (Report No. 28). Philadelphia, PA: The Annenberg Public Policy Center of the University of Pennsylvania.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370.

Zins, J. E., & Elias, M. J. (2006). Social and emotional learning. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 1–13). Bethesda, MD: National Association of School Psychologists.