ELSEVIER
SAUNDERS

Adolesc Med 16 (2005) 427–446

ADOLESCENT
MEDICINE
CLINICS

Media Ratings for Movies, Music, Video Games, and Television: a Review of the Research and Recommendations for Improvements

Douglas A. Gentile, PhD[a,b,c,*], Jeremy Humphrey, BS[a], David A. Walsh, PhD[c]

[a]*Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011-3180, USA*
[b]*Institute of Science and Society, Iowa State University, Ames, IA 50011-1302, USA*
[c]*National Institute on Media and the Family, 606 24th Avenue South, Suite 606, Minneapolis, MN 55454, USA*

The "nature-nurture" debate has intrigued psychologists since the time of Plato. As modern experimental and behavior genetics techniques have improved, it has become clear that environmental influences are critical for the development of most personality traits and behaviors [1]. In the past 60 years, the environment in which children and adolescents develop has changed in ways that never have been seen before in human history—it has become media-saturated. There are far more types of media and children have far greater access to those different types of media than ever before. Therefore, it is possible for youth to observe attitudes, beliefs, and behaviors modeled daily that are different from those that are observed in the child's natural environment. It has not gone unnoticed by parents, politicians, and (more importantly, perhaps) researchers that this new "environment" may have profound effects on child development.

Although attention to media influences on children and adolescents often is intensified by tragedies, such as the spate of school shootings in the United States

Please do not cite or copy without permission. Dr Walsh has a proprietary interest in the KidScore rating system that is described in the article.

* Corresponding author. Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, IA 50011-3180.
E-mail address: dgentile@iastate.edu (D.A. Gentile).

1547-3368/05/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.admecli.2005.02.006

in the late 1990s, researchers have been examining the effects for almost 50 years. It is now known that media can be a risk factor or protective factor for healthy development, depending on the amount and content of media consumed [2]. For example, watching a great deal of entertainment television has negative effects on school performance [3], whereas watching educational television programs, such as *Sesame Street*, has positive effects that can still be observed in adolescents' school performance [4]. Hundreds, if not thousands, of studies have been conducted on the multiple effects that the various media can have.

As the first medium to receive attention, television (and film, in general) has the largest body of research to draw upon. Bandura's famous "Bobo doll" studies provided evidence that humans can learn aggressive behaviors from film [5], especially if the behavior shown was not punished, and even if it was [6]. Numerous experimental, correlational, and longitudinal studies on violent media have confirmed that media violence leads to increased aggressive thoughts, feelings, and behaviors (see [7] for a review). This relation is extremely robust, and crosses several media, including television, movies, video games, and to a lesser extent, music. Although there is less research on the effects of sexual media on children, it is still remarkably consistent in that it does seem to affect children's attitudes about sex [8].

Overall, the accumulated research on the effects of media on children is comprehensive and clear; parents are correct to be concerned about the media environment that surrounds their children. Thus, the need for a competent rating system for the media that is available to the public is a valid one. Individuals, especially parents, should be provided with appropriate, reliable, and valid information that can allow them to make informed decisions regarding the media consumption of their household. This is the goal for media ratings systems. Most major media (TV, movies, music, video games) now have ratings systems. Although each industry has its own system (sometimes more than one!), several themes are common. They all were instituted under political pressure, and they all suffer from serious flaws. The history of the major ratings systems are reviewed, and is followed by a discussion of how they could be improved from a scientific standpoint.

History and current configurations of the ratings systems

The authors have taken the liberty to describe the research on the effects of ratings on all ages of children for two reasons: the consequences of what children view often are played out during adolescence, and most improvements to the ratings would be made best at a holistic system level.

Movies

The movie industry was the first to adopt a rating system. In 1968, The Classification and Ratings Administration (CARA) was created by the

Motion Picture Association of America (MPAA; the trade association that is funded by the major movie studios) to provide the public with content information of films. This system is voluntary, although CARA was created in response to the threat that the government might take political action and impose viewing restrictions on minors. The goal of the movie rating system is to provide parents advance knowledge about movies that will permit them to make a good informed decision as to whether their child should see the movie.

Originally, movies were assigned a rating of G for "General Audiences," M for "Mature Audiences," R for "Restricted under 16, need parent or guardian present," or X for "No admission under 17." These ratings have undergone many changes over the years. Currently, there are five different ratings that a movie can be given (Fig. 1).

All of the ratings attempt to gauge, at least approximately, the "proper" age a viewer should be to watch that particular movie. Because research has shown that content-based ratings systems are preferred to age-based ratings (see later discussion), the movie ratings now often include limited information about why a movie was given its rating (eg, language, sexual content, violence) along with the age-based rating, although it is provided in substantially smaller print than the age-based rating.

Movie Ratings	
Rating	Explanation
GA	General Audiences – All ages admitted
PG	Parental guidance suggested. Some material may not be suitable for children
PG-13	Parents strongly cautioned. Some material may be inappropriate for children under 13
R	Restricted. Under 17 requires accompanying parent or adult guardian
NC-17	No one under 17 admitted
Music Rating	
PARENTAL ADVISORY EXPLICIT CONTENT	The recording may contain strong language or depictions of violence, sex or substance abuse. Parental discretion is advised.

Fig. 1. The major ratings systems and their meanings. (Courtesy of the Motion Picture Association of America; with permission.)

Video Game Ratings According to the Entertainment Software Rating Board	
EC	Titles rated **EC - Early Childhood** have content that may be suitable for ages 3 and older. Contains no material that parents would find inappropriate.
E	Titles rated **E - Everyone** have content that may be suitable for persons ages 6 and older. Titles in this category may contain minimal violence, some comic mischief and/or mild language
E^+	Titles rated **E+ - (Everyone 10 and older)** have content that may be suitable for ages 10 and older. Titles in this category may contain more cartoon, fantasy or mild violence, mild language, and/or minimal suggestive themes.
T	Titles rated **T - Teen** have content that may be suitable for persons ages 13 and older. May contain violent content, mild or strong language, and/or suggestive themes.
M	Titles rated **M - Mature** have content that may be suitable for persons ages 17 and older. Titles in this category may contain mature sexual themes, more intense violence and/or strong language.
AO	Titles rated **AO - Adults Only** have content suitable only for adults. Titles in this category may include graphic depictions of sex and/or violence. Adult Only products are not intended for persons under the age of 18.

Television Ratings According to the TV Parental Guidelines Monitoring Board	
TV Y	This program is designed to be appropriate for all children. Whether animated or live-action, the themes and elements in this program are specifically designed for a very young audience, including children from ages 2 - 6. This program is not expected to frighten younger children.
TV Y7	This program is designed for children age 7 and above. It may be more appropriate for children who have acquired the developmental skills needed to distinguish between make-believe and reality. Themes and elements in this program may include mild fantasy violence or comedic violence, or may frighten children under the age of 7. Therefore, parents may wish to consider the suitability of this program for their very young children.
TV Y7 FV	For those programs where fantasy violence may be more intense or more combative than other programs in this category, such programs will be designated TV-Y7-FV.
TV G	General Audiences
TV PG	Parental guidance suggested
TV 14	Parents strongly cautioned
TV MA	Mature audience only
V	Includes violence
S	Includes sexual situations
L	Includes coarse language
D	Includes sexually suggestive dialogue

Fig. 1 (*continued*).

Producers or distributors submit their movies for review under their own free will. However, in essence, they are required to do this because the National Association of Theater Owners will not show unrated movies. The CARA ratings board is composed of up to 12 individuals who are not supposed to be connected to the movie industry, and should be parents. Board members also must "be possessed of an intelligent maturity, and most of all, have the capacity to put

themselves in the role of most American parents." There are 9 junior raters, whose hours range from full time to part time. These raters may serve for a total of 5 years. There also are 3 senior raters, who have special responsibilities, such as leading discussions about ratings with directors. Senior raters have no restrictions on the number of years for which they can serve. One former rater broke the MPAA's secrecy restrictions to comment that there are no clear standards for how ratings are provided (ie, what distinguishes R from PG-13) [9]. Additionally, raters are required to give feedback on ways to reduce a rating if the director desires (eg, an R rating on a film that initially received an NC-17). The identities of all members are kept confidential. There is also a board chair, whose identity is public knowledge. In case of a tie between the 12 raters, the board chair casts the deciding vote [10–12].

Music

The music industry voluntarily began to place parental advisory labels on albums with potentially objectionable lyrics in 1985. These warning labels received the nickname "Tipper Sticker" because one of the founders of the Parents Music Resource Center, Tipper Gore, implied at congressional hearings that mandatory regulation would be required if the industry did not self-regulate. The warning labels make no recommendation regarding age or specifically what type of explicit content the album contains. All labels give the vague statement: "Parental Advisory Explicit Content" (see Fig. 1). The choice to label an album is determined solely by the record companies and artists, and there are no standards for determining which recordings should be labeled. Any restrictions on the sale of these marked albums are decided by individual stores [10,12].

Video games

Similar to other media, a rating system for video games and software was developed in response to government pressure in 1994. Two rating systems emerged—one sponsored by the Interactive Digital Software Association (IDSA), and one developed by the Software Publishers Association (SPA).

The SPA created the Recreational Software Advisory Council (RSAC), which decided to classify their games by way of content-based statements. A questionnaire that uses standardized definitions to determine the amount of content on a scale of 0 to 4 for the topics of sex, violence, and language in a game is completed by the game developers.

The IDSA developed the Entertainment Software Ratings Board (ESRB), which decided to classify their games into age-based categories that are similar to movie ratings. There were originally four categories: K-A (Kid-Adult), Teen (Ages 13+), Mature (Ages 17+), and Adults Only (Ages 18+). Modifications

were soon made to split the K-A category into Early Childhood (ages 3 and older) and Everyone (Everyone, suitable for ages 6+). The E category was further modified in 2005 with the addition of the E10+ rating (Everyone 10 and older) (see Fig. 1). Content descriptors were added to provide a general idea to the consumer about the levels of violence, sex, objectionable language, and drug usage that are contained in the games. The ESRB beat the RSAC in the battle of which rating system would be prominent, partly because the ESRB is funded by the video game industry. The ESRB now provides virtually all of the ratings for console games and computer games, whereas the RSAC provides ratings for Internet sites [10,13]. To add to the confusion, the coin-operated video game industry has an entirely different rating system [10].

The game manufacturer begins the rating process by submitting game footage and any other information to the ESRB. Three trained raters review the material in the absence of the others. These raters are not permitted to have any ties to the video game and computer industry, and their names are kept confidential. If two of the raters apply the same rating, that rating is used [12]. When there is dissonance between the ratings, additional raters are recruited to reach a "broader consensus" [14] after which the final rating is approved and released.

Television

Until 1997, television had no rating system. Occasionally, a network would provide the message that "viewer discretion is advised." The Telecommunications Act of 1996 mandated that televisions include a "v-chip," which would allow programs with particular content to be blocked. The television industry was given 1 year to develop a rating system "voluntarily," after which point the Federal Communications Commission would appoint a committee to create one [15].

Jack Valenti, architect of the movie ratings system, helped to create a similar age-based rating system for TV in 1997. It was denounced immediately by researchers, parents, and child advocacy organizations because it gave no information about the content of shows and could backfire by attracting younger children to shows that were rated as being intended for older audiences (the "forbidden fruit" effect) [16]. In response, the ratings were amended to include content-specific categories as a complement to the age-based categories. The six age-based and six content-based ratings are shown (see Fig. 1). For shows that are rated TV-Y, TV-Y7, or TV-G, no content ratings are provided (with the exception of FV, which is only used with TV-Y7). For programs that are rated TV-PG or greater, content descriptors may be used to indicate what type of objectionable material that program contains (although NBC refuses to provide content information).

Each network is responsible for rating its own programming. The network provides a "prerating" for each show before shipping it to a local affiliate [15].

This "prerating" can be changed by the local broadcaster in the case of a conflicting opinion, but most do not review the program again and simply accept the "prerating" to save time and manpower. The possibility exists for two networks to give different ratings for the same program (and many cable channels have their own ratings, which can cause even more confusion). There also are no industry standards by which programs are rated. Given these factors (eg, lack of independence, unclear standards,), it is critical to test the reliability and validity of the rating systems.

The goals of ratings systems, and research findings related to those goals

A useful rating system should provide as much relevant and correct information to parents and caregivers as possible, in as simple and understandable a format as possible [17]. In turn, this will allow parents to make informed choices about what is appropriate for their children. The goals are not to censor or dictate taste, but to provide valid objective descriptive information [17]. From a public health perspective, the ultimate goals are to empower parents, and thus, decrease the risk of harm to children and adolescents.

For these goals to be met, several conditions must exist. First, it must have been demonstrated that a credible risk of harm exists, and that the conscientious use of ratings can lower that risk. As an example, playing violent video games has been shown to cause increases in aggressive behaviors [7]; however, if parents use the video game ratings to help decide what games their children may rent or buy, children engage in fewer aggressive behaviors [18]. Therefore, ratings systems do have the potential to meet the ultimate goal of decreasing the harm of inappropriate media on children and adolescents; however, this potential could be undermined completely if the ratings are not reliable, valid, easily understandable, or easily accessible. Furthermore, it also could be undermined if the ratings make children more interested or likely to view/hear/play a media product (the "forbidden fruit" hypothesis). Each of these issues is addressed below.

Reliability and validity

With respect to reliability and validity, it is critical to know that these terms have definite scientific and statistical meanings (ie, they are measurable and testable). There are several types of reliability and validity, and there is not sufficient space to discuss them all here (see Hartmann and George [19] for an overview). Among the several aspects of reliability, three are especially important here: interrater reliability, consistency, and temporal stability. Interrater reliability is demonstrated by having several raters all see the same thing—if one judge rates a show as having a lot of violence, the other judges would rate it similarly. This demonstrates that the judgments are not idiosyncratic to one rater. For high

interrater reliability, the variables that are being rated must be well-defined. The consistency aspect of reliability means that if two movies had empirically identical content, they should be given the same ratings. For example, if they both included violent content, they should be labeled with a violence descriptor (if the rating system includes content descriptors). The temporal stability aspect of reliability means: (1) that the same show would be given the same rating if rated twice, and (2) that the rating criteria should not shift across time. All of the ratings systems fail several of these aspects of reliability (see later discussion).

Reliability is not only important on its own, but also because it has grave implications for validity. If a rating is not reliable, it cannot be valid (ie, reliability is a necessary, but not sufficient, condition for validity). As an example, if a clock is always 5 minutes fast, is it reliable? Yes, but it also is consistently wrong (and therefore invalid). A measurement is valid if it accurately measures what it is supposed to measure. As with reliability, there are several types of validity, three of which are of particularly germane: content validity, construct validity, and criterion validity.

Content validity refers to whether a rating system measures what it claims to measure. For example, if a system was intended to rate the amount of physical violence in a video game, but it did not consider shooting to be physical violence, or it included verbal aggression, the rating would not have content validity. Content validity forces researchers to define clearly what they are intending to measure, and typically requires a panel of content experts to ensure that the definitions are correct. In the case of media ratings, parents would be one obvious panel of experts (ie, if parents do not agree that the ratings measure what they expected them to measure, then the ratings would not be valid).

Construct validity refers to the agreement between a measurement rating and the theoretic construct that is intended to be measured. Violence is an example of a construct. Because it can be defined differently by different people, a clear definition is required. Then the construct validity can be determined by comparing whether the measurement of violence is correlated with other theoretically relevant variables, such as trait hostility (this is convergent validity), and if the measurement is not correlated with theoretically nonrelevant variables, such as height (this is divergent validity).

Criterion validity refers to the agreement between a measurement rating and other ratings that already were proven to be valid. Together, these aspects of reliability and validity are the hurdles that media ratings systems are attempting to clear. Jack Valenti (in his role as MPAA CEO) wrote a letter the authors that stated, "In brief, there is no 'validity research' available in the social scientist lexicon." This quote makes it clear that Mr. Valenti does not understand what scientists mean by validity or how they can provide evidence (or a lack thereof) of validity.

Interrater reliability

To our knowledge, none of the media ratings boards publishes statistics on interrater reliability. This highlights an additional problem with media ratings as

they currently exist—they are opaque. We do not know how raters are trained, how categories are defined, or whether interrater reliability ever is achieved. The ratings process should be transparent and subject to scientific scrutiny.

Consistency

Over the years, there have been consistent complaints about the lack of consistency in many media ratings [20,21]. Examples of ratings that seem to inconsistent abound. For example, one commentator noted: "The real villain here is the ...MPAA. Its 'Kangaroo Jack' ruling [it received a PG despite being considered "profoundly unsuitable" for kids] is another nail in the coffin for a ratings board that has shown itself to be wildly out of touch with parental standards and totally capricious in its judgments, having given an R rating, for example, to 'Billy Elliot,' preventing scores of kids from seeing a truly inspirational movie" [22].

To our knowledge, there has been no formal research conducted on the subject of the reliability and validity of music ratings. The same parental advisory warning sticker is used for albums that contain a few instances of objectionable content, and others that may contain hundreds of instances of objectionable content. For example Eminem's album "The Marshall Mathers LP" contains hundreds of swear words, and many blatant violent, sexual, and drug references. In contrast, American HI FI's album (self-titled) contains four swear words, and some sexual references. Under this system, a parent could check the lyrics of an album with a warning sticker that has little objectionable content and decide it is appropriate for the child or adolescent. Then, that parent may generalize future judgments about albums with warning stickers on that review of that album, even though the amount of content that is contained in two albums with the same warning label can be drastically different.

With regard to video games, studies show that almost half of E-rated games that include violence did not receive a content descriptor for violence [23], and almost half of T-rated games included content (eg, violence, sexual images, drug use) that was not listed in the ESRB-provided content descriptions [24]. Similarly, the TV ratings do not apply the content descriptors consistently. In one content analysis, more than three fourths of shows with violence and nine out of ten shows with sex did not receive the V or S content descriptors [16]. Among children's shows, 81% that included violence did not receive the FV content descriptor. Clearly, the ratings have a long way to go to become consistent.

Temporal stability

One issue that all ratings systems face is that parents need to learn what each of the symbols mean. If the meaning of the ratings changes, then parents cannot rely on them. For example, once parents learn what types of content can exist in a PG-13-rated movie, they probably do not expect it to change. Yet, an increasing number of studies are showing what is known as "ratings creep." The idea of ratings creep simply suggests that increasingly more objectionable content

is trickling down to less restrictive ratings. For example, the *Alien* series and *Predator* series of movies were given R ratings, starting in 1979; however, the recently-released movie *Alien Vs. Predator* was given a PG-13 rating. It is highly unlikely that this newly-released movie has less violence and adult language than its predecessors.

An examination of violence in G-rated animated films found that, over time, there is a statistically significant increase in the duration of violence within the films, even when controlling for increased film length [25]. This is especially important because animated G-rated films are designed especially for young children. In an analysis of 1906 movies that were released between 1992 and 2003, researchers found significantly higher objectionable content in movies as a function of time; this suggests that the MPAA applied less stringency in its age-based ratings over time [26]. In fact, the Federal Trade Commission stated in 2000 [12]:

> According to the MPAA, PG-13 films have no rough or persistent violence, and the existence of such violence will cause a film to be rated R. Nevertheless, the Commission's study of the ratings explanations indicates that CARA often describes the violence in PG-13 films in terms synonymous with rough and persistent violence, for example, as "intense," "strong," "disturbing," "brutal," "graphic," "shocking," "non-stop," and "pervasive." These words are identical or similar to words often used to describe violence in R-rated movies, causing some to question the usefulness of the ratings for helping parents distinguish the amount and kind of violence in PG-13 films from that in R-rated films.

Parents would be most likely to notice ratings creep in the categories that are closest to the most restrictive category (eg, in PG-13 movies, TV-14 shows, and T games—the adolescent categories), as previously restricted content becomes more "acceptable." In an analysis that compared validated parent ratings with movie, television, and video game ratings, this is exactly the pattern that is seen [27]. The greatest disagreement with the ratings tends to be in these categories, which suggests that mature content is filtering down to the less restrictive ratings. In a study of films between 1992 and 2003, researchers found that a movie that is rated PG-13 today is approaching the average amount of violence, nudity, and foul language of R-ratings a decade ago [26].

Taken together, the studies that address the reliability of media ratings show that there is a grave lack of reliability in how/when the ratings are assigned. Several studies have attempted to go beyond the issue of reliability to measure the validity of media ratings; reliability is a necessary precondition for validity. Without strong reliability the ratings cannot be valid.

Content validity

For reliability and validity to be possible, clear definitions of what is being measured by the ratings need to be created. To the authors' knowledge, no media ratings systems have clear guidelines and definitions of important vari-

ables, such as violence. This is odd, because reliable approaches have been created by researchers to conduct content analyses.

Regardless of the lack of definitions, content validity can be measured by comparing the existing ratings with panels of experts. Two types of experts seem particularly relevant when considering the validity of ratings: parents (who are the users of ratings) and child development experts (who can judge based on what is likely to be beneficial or harmful to children). Although there is not a great deal of research, the small body that exists suggests that the ratings are considerably less valid than one would hope.

In their analysis of 1269 movies that were released between 1992 and 2003, Thompson and Yokota [27] compared the MPAA ratings with two independent content-based ratings systems that are designed to provide better information to parents (Screen-It! [28] and Kids-in-Mind [29]). Although these two content-based systems correlated highly with each other (r = 0.83), there was much greater variability when comparing the content-based systems with the MPAA age-based rating and even with the MPAA's content descriptors. In a survey of parents' use of the television ratings, fully half say they have disagreed with a rating that a show was given [15]. In a study of parents' perceptions of video game ratings, there was high agreement when games were obviously violent or nonviolent, but considerable disagreement when games included cartoon violence or fantasy violence [10].

In a notable study of 1332 TV shows, the investigators coded the shows on dimensions that pose the greatest degree of risk for harmful effects on children and adolescents, and compared these with the TV ratings [30]. Based on previous research, a great deal is known about what types of portrayals of violence and sex are most likely to result in negative effects. For example, not all portrayals of violence result in a greater likelihood of youth copying the violence. If the violence appears in the context of an antiviolence theme, it was not considered to be a high-risk portrayal. More than two thirds (69%) of children's shows with high-risk violent content was rated as TV-Y and did not include the "V" content descriptor. Among general audience shows (not specifically designed for children) that included high-risk violent content, 40% were rated TV-PG, and 65% had no "V" content descriptor. Among general audience shows that included high-risk sexual content, 29% were rated TV-PG and 80% did not include the "S" content descriptor.

Construct validity

The studies that were described above also provide evidence for a lack of sufficient construct validity in media ratings. Construct validity can be measured by looking for high correlations between theoretically relevant measures, such as a relation between content-based and age-based ratings. In general, there is a positive correlation between independent ratings and industry-provided ratings (ie, when movies receive more restrictive ratings [PG, PG-13, R] they also receive higher scores for violence, sex, and profanity as measured by content-

based ratings) [27]. There is a great deal of variability, however, and it does not seem to be random. When there are disagreements between industry-provided ratings and parents, content analyses, or other ratings systems, they almost always occur because the industry-provided ratings did not include information about content. From the perspective of the parent, this is likely to mean two things: (1) if a media product is labeled as not for children (R, M, TV-MA, or Parent Advisory labels), parents almost always agree; and (2) many times, the parents will find content in less restrictively labeled products that they did not expect to find, and therefore, will find the ratings too lenient.

Criterion validity

To the authors' knowledge, only one study clearly addressed the criterion validity of media ratings. To measure criterion validity, the results of a rating system is compared with another that previously was shown to be valid. The National Institute on Media and the Family created a combination age-based and content-based ratings system that could be applied to any media products. This system (called KidScore) was validated nationally with a sample of 600 randomly selected households. Walsh and Gentile [27] compared the industry ratings with the validated KidScore ratings for 276 movies, 253 TV shows, and 166 video games. If the industry ratings were valid, then logically they should match up with the KidScore ratings. The results showed the same pattern as other research—that parents always agree when a product is given an R, M, or TV-MA rating, but there is far less agreement with other ratings.

For movies, Walsh and Gentile [27] found that parents considered only 50% of G-rated films as completely appropriate for children ages 3 to 7. Similarly, 63% of viewed PG-rated films were deemed clearly appropriate for children ages 8 to 12, and 60% of viewed PG-13 rated films were considered appropriate for teens ages 13 to 17. (Across these three rating categories, this results in an average inconsistency of 42%.) Given the difference between the industry's recommendation of viewing age and the agreement of parents, the reliability and validity of movie ratings are low. CARA is providing the service that they claim as their most important goal only 58% of the time for films that are rated G, PG, and PG-13. By most educational standards, they are failing.

For video games, Walsh and Gentile [27] found that parents rated 67% of viewed E-rated video games as clearly appropriate for children ages 3 to 7, and 87% of the same games as appropriate for children ages 8 to 12. In contrast, of the games that were rated T for Teen, only 43% were deemed appropriate for teens ages 13 to 17. This large discrepancy may be accounted for, in part, by the fact that 48% of T-rated games did not receive an ESRB content descriptor for which content in the game was observed [24]; however, the raters may not be completely at fault. Recall that the game manufacturer submits a video of game footage along with any other information that should be considered. Especially if a game is marketed for consumers ages 13 to 17, it would be in the manufacturer's best interest to not include any content which may be grounds for

a M rating. Raters cannot rate what they do not observe. The research suggests that the industry is providing fairly reliable and accurate ratings for E-rated and M-rated games, with a weakness in identifying fantasy violence as violence [23]. The T rating seems to be the least valid, which may be additional evidence of ratings creep.

Criterion validity was lowest for television. Three quarters (76%) of shows that were rated TV-Y were considered to be clearly appropriate for children ages 3 to 7; however, this seems to be the most consistent rating. TV-G ratings, which are rated appropriate for all ages, although not directed specifically at children, were only judged as clearly appropriate for children ages 3 to 7 only 40% of the time. Fifty-seven percent of TV-Y7 programs were rated clearly appropriate for children ages 8 to 12; TV-PG programs were considered appropriate for 8- to 12-year-old children only 23% of the time. Finally, only 15% of television shows that were rated TV-14 were rated clearly appropriate for the adolescents aged 13 to 17 (again suggesting the issue of ratings creep). If the television show ratings were completely valid, the above percentages should be near 100% each.

Taken as a whole, the research on the reliability and validity of current media ratings systems demonstrates that there is too little of each. Parents can rely only on the most restrictive ratings. When an industry states that a particular show, film, or game is not for children, parents almost always agree. But beyond that, there is considerable disagreement. The shame is that it does not need to be this way. There is a strong body of research that shows what parents want in a rating system, and what makes a good rating system.

Research on what makes a good rating system

Beyond the empiric issues of reliability and validity, there are some practical issues to consider about the ratings systems. Parents are the intended end-users of the ratings, so they should be understandable and easy for parents to use. Almost all parents (90%) agree that media ratings are a good idea [15]. Yet, ratings are not used as widely as one might expect, although the pattern of use reflects how long they have been in existence. Movie ratings are the oldest, and 78% of parents say that they have used them to guide their family's media choices. Only approximately half of parents have ever used the music advisories (54%), video game ratings (52%), or TV ratings (50%) [31]. Approximately half or fewer believe that the ratings are "very useful" (45% for movie ratings, 48% music, 53% video games, and 38% for TV) [31]. Given that parents want ratings, why do so few find them useful? One reason is that parents are able to notice the lack of reliability and validity themselves. For example, only 2% of parents who have used the TV ratings system believe that they always reflect the content of the shows accurately [31]. Furthermore, the ratings as they currently exist are confusing. For example, fewer than half of parents (43%) say that they under-

stand all of the symbols that are used in the TV rating system [32], and when pressed to name what the symbols mean, correct responding decreased to as low as 4% [31]. Each medium has different symbols from the others, which creates what some have likened to a bad Scrabble hand of confusing letters and symbols.

When parents are asked about what they would like in a ratings system, they overwhelmingly state that they would prefer content-based ratings, rather than a summary age-based rating [33–36]. Yet, the media industries have been reluctant to provide this information, and only have included it as an extra piece of information that is tacked onto the existing rating. This is confusing, given that it is not only parents who would prefer content-based ratings.

When it was announced in 1996 that there would be a television rating system, Children Now convened a meeting of 18 child development and media experts to compile recommendations for the new system [17]. The experts overwhelmingly preferred a system that combines content- and age-based ratings. The experts stressed the importance of rating content for the categories of violence, sex, and language. Including content descriptors in the rating system is important because parents have differing levels of concern for different media content. Parents of children who are ages 3 to 7 are only moderately concerned with sex and language, but are concerned highly about violence when their child is a boy, and concerned highly about fright when their child is a girl [33]. When children reach the ages of 8 to 12, topics of sex become a greater concern for parents, in general. Parents of boys continue to be more concerned about violent content and risk-taking behaviors than parents of girls in this age group. For parents of teenagers, sex was reported as the greatest concern, with violence, language, and risk-taking as areas of moderate concern. Overall, more than three quarters of parents say it is "very important" to know the amount of violence, vulgar language, and sexual content that are present in media [35].

Beyond parent preferences, research has documented another reason why content-based rating systems should be favored over age-based ratings: the "forbidden fruit" effect, which proposes that certain media presentations may be perceived as more desirable because they have been labeled with an age-restrictive rating, implying that the child is "too young" to view the media. Researchers have found empiric evidence for the forbidden fruit effect [34,36,37]. In one study, older children and boys were more interested in an identical video clip that was given a rating of PG-13 or R. Specifically, not a single older boy (ages 10–14) in the sample selected the movie when it was rated G, but when the movie was rated PG-13 and R, 53% and 50%, respectively, wanted to see the movie. Also, the statement, "parental discretion advised" invoked the forbidden fruit effect in the group of older boys, whereas the statement "viewer discretion advised" had no effect on any group, and actually caused children, specifically younger girls (ages 5–9), to be less interested in the program, a "tainted fruit effect." An additional experiment the following year yielded similar results. Again, MPAA ratings of PG-13 and R increased children's desire to see the movie, and a rating of G decreased this desire. Content descriptors for violence

were added to the design as an independent variable to the second experiment. Violent content descriptors did not increase a child's desire to view a program, provided no age-based rating accompanied the content descriptor. A recent meta-analysis of research on ratings found that across studies, age-based ratings consistently increased interest in children and adolescents starting at about age 8 and continuing up to about age 22 [36].

Summary and recommendations for improvements

The research on the various effects of media on children is broad, deep, and compelling. Almost all pediatricians believe that media use affects children on several important dimensions, including school performance, eating habits, overweight/obesity, aggression, high-risk behaviors, and sexual behaviors [38]. For this reason, it is important to provide clear, reliable, and valid information to parents and caregivers to help them choose what media their children may watch/hear/play.[1] Furthermore, parents clearly want the information. Unfortunately, as this article shows, parents are hindered by a lack of reliability and validity in the ratings. What this means in practice is that even if a parent is carefully and conscientiously using the ratings, his/her children will continue to be exposed to inappropriate material. We are forced to conclude that nothing short of a complete overhaul of the media ratings will solve the problems that are inherent in the current systems. Our recommendations include long-term and short-term measures to address the flaws in the system.

Long-term recommendations

Universal rating system

Our strongest recommendation is for the development of a single universal system for rating media products [32]. The current alphabet soup of systems is confusing for parents. More importantly, multiple ratings systems are more complicated to test, validate, and monitor than a single system. One universal system would be far more efficient to train raters and parents to use.

Many industry representatives have argued that a universal system is not possible, and that because the various media are different, the ratings also must be different [39–42]. Both of these claims are difficult to support. For example, the KidScore and Screen-It! ratings systems have been used since 1996 to rate movies, TV shows, and video games. Furthermore, although the authors agree that movies, television, music, and video games are different in important ways, the concerns that parents and physicians have about violence, sexual content, and

[1] The research on music is far less compelling than that on screen-based media. See Gentile [7] for a review of the media violence literature covering several media.

language are similar across all media types. In fact, the various media ratings systems already measure essentially the same things. It is time to make use of the accumulated research data to make major improvements across all ratings systems. The authors first lobbied for this change in 2001. Since then, several others have joined them [16,36]. Even the Directors Guild of America has joined the call for a universal ratings system [43]. Perhaps, more importantly, in a national survey of households that the authors conducted in 2001, 84% of parents said that they would support the creation of one standard rating system that could be used by all media [44].

It has been argued that one company (eg, a video game company) could not rate movies or TV shows or music. That claim seems unlikely to be true when one considers the great deal of vertical integration that already exists. Only five companies own most of the movie production companies, movie distribution networks, TV networks, record labels, video game companies, newspapers, and magazines. Single companies already use or support several types of raters for different types of media. It actually would be less expensive and more efficient if they only needed to have one type of rater.

Another common criticism of the proposal for universal ratings is that parents like the current systems [42]. The authors do not dispute this. Parents want ratings and they like having some information more than none. But the authors contend that they would like better ratings even more.

System design

The research that was reviewed above provides several clues about how to design a better ratings system. Therefore, appropriate academic and medical experts should work together with media representatives to create a universal system, with opportunities for public comment. To be truly effective, a universal system should include the following characteristics:

Explicit goals
Explicit guidelines for raters, with clear definitions of relevant constructs
Raters should be independent of each other, and independent of the media industry
High interrater reliability
A format that is easy for users to understand
Appropriate age-based information
Appropriate content-based information, including ratings based on scientific information about the effects of media on child development. Currently, the ratings are based solely on what is likely to be offensive to parents, rather than on what is likely to be harmful to children and adolescents [17,45]. This would include a change to how fantasy or cartoon violence is rated [10].
The content information should no longer be described using euphemisms, but should describe the content clearly [34].

High reliability of ratings across media products
High reliability of ratings across time
Demonstrable validity of the ratings (eg, concordance with parent ratings)
Coverage of a broad range of products (ie, TV, movies, and video games at a minimum)
Products rated in a timely manner
Broad and easily accessible distribution

Authority

Since their inception, the ratings have been implemented on a voluntary basis. The combination of industry and public pressure to use ratings has encouraged the makers of media products to use the ratings to label their products; however, the economic stakes of ratings also play a role. The competition for the public's eye and money has increased sharply. The economic temptations to "downrate" a product to capture a larger audience have increased [22], while at the same time producers continue to outdo the previous season by including more "edgy" material—more violence, more sexual situations, and more adult language. This is done partly because the public becomes desensitized. It also is seen in the ratings creep.

It is time to create an independent ratings oversight committee. At a minimum, this committee would conduct regular research on the reliability and validity of media ratings to maintain standards across media and across time. Additionally, it could have the authority to assign or approve ratings. This committee could be composed of industry representatives, child development experts, physicians, and psychometricians. It would be best, however, if it were independent of any one medium. The authors make no recommendation about whether this oversight committee should be created by the media companies, whether it should be an independent organization, or how it should be funded. It has been argued that it would not be possible for a single organization to rate all of the products. If this is true (and it is unclear that it is), then the oversight committee should be responsible for retraining the raters who already are doing the ratings. It would represent few additional resources to have all of the existing raters trained, and then have continuing training to keep their ratings consistent across time.

Short-term recommendations

The authors recognize that long-term solutions may take many months or years to implement. In the interim, parents and physicians need short-term tactics to guide them in using the current imperfect ratings systems.

Education

The public and the medical community must continue to be educated about the ratings systems and why they should be used. Educational efforts should include awareness of the existing systems and the ratings in each. At the same

time, parents and physicians should be made aware of the problems with the current systems, and the dialog surrounding the evolution of these systems. They should realize that a rating is not a "seal of approval," and that they should continue to monitor their children's and adolescents' use of media products, even if the rating suggests that the products are age-appropriate. Perhaps more importantly, educational efforts should seek to involve parents and physicians in the process of reinventing the ratings as outlined above.

Research

As shown in this article, research has enabled us to understand how reliable and valid the current ratings are, what parents want in a ratings system, and how to create better ratings. There is still a great deal of research to be done, however, including more investigation of the effects of sexual content and adult language on children's and adolescents' health and well-being. Furthermore, the efficacy of ratings continues to need to be addressed.

The health care community already has expended a great deal of time, expertise, and energy advocating for strategies that would reduce the harmful effects of media on youth, as well as to enhance the positive effects that media can have on them. The success of those efforts is impeded if the ratings systems are unreliable or invalid. With continued questionable reliability and validity, one of two results will occur. Either parents will realize that the ratings are unreliable and will cease to use them, or they will continue to use them but children and adolescents will continue to be exposed to inappropriate material. Under either scenario, the goal of the entire effort is undermined and youth will remain at risk.

References

[1] Plomin R, DeFries JC, McClearn GE, et al. Behavioral genetics. 3rd edition. New York: Freeman; 1997.

[2] Gentile DA, Sesma A. Developmental approaches to understanding media effects on individuals. In: Gentile DA, editor. Media violence and children. Westport (CT): Praeger; 2003. p. 19–37.

[3] Huston AC, Donnerstein E, Fairchild H, et al. Big world, small screen: the role of television in American society. Lincoln (NE): University of Nebraska Press; 1992.

[4] Anderson DR, Huston AC, Schmitt KL, et al. Early childhood television viewing and adolescent behavior: the recontact study. Monographs of the Society for Research in Child Development 2001;66(1):1–147 [Serial No. 264].

[5] Bandura A, Ross D, Ross SA. Transmission of aggression through imitation of aggressive models. J Abnorm Soc Psychol 1961;63:575–82.

[6] Bandura A, Ross D, Ross SA. Vicarious reinforcement and imitative learning. J Abnorm Soc Psychol 1963;67:601–7.

[7] Gentile DA, editor. Media violence and children. Westport (CT): Praeger; 2003.

[8] Brown JD, Steele JR, Walsh-Childers K, editors. Sexual teens, sexual media. Mahwah (NJ): Lawrence Erlbaum Associates; 2002.

[9] Waxman S. Rated S, for secret. Washington Post April 8, 2001;G01.

[10] Funk JB, Flores G, Buchman DD, et al. Rating electronic games: violence is in the eye of the beholder. Youth Soc 1999;30:283–312.
[11] Cantor J. Rating systems for media. In: Johnston DH, editor. Encyclopedia of International Media and Communications. San Diego (CA): Elsevier Science; 2003. p. 47–57.
[12] Federal Trade Commission. Marketing violent entertainment to children: a review of self-regulation and industry practices in the motion picture, music recording & electronic game industries. Available at http://www.ftc.gov/reports/violence/vioreport.pdf. Accessed April 12, 2005.
[13] Cesarone B. Video games: Research, ratings, recommendations. Eric Digest Report #EDO-PS-11; 1998. Available at: http://ceep.crc.uiuc.edu/eecearchive/digests/1998/cesar98.html. Accessed April 12, 2005.
[14] ESRB game ratings page. Available at: http://www.esrb.org/esrbratings.asp. Accessed December 11, 2004.
[15] Kunkel D, Farinola WM, Cope KM, et al. Rating the TV ratings: one year out: an assessment of the television industry's use of the V-chip ratings. Menlo Park (CA): Kaiser Family Foundation; 1998.
[16] Kunkel D. The road to the v–chip: television violence and public policy. In: Gentile DA, editor. Media violence and children. Westport (CT): Praeger; 2003. p. 227–46.
[17] Children Now. Making television ratings work for children and families: The perspective of children's experts. Oakland (CA): Children Now; 1996.
[18] Gentile DA, Lynch PJ, Linder JR, et al. The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. J Adolesc 2004;27:5–22.
[19] Hartmann DP, George TP. Design, measurement, and analysis in developmental research. In: Bornstein MH, Lamb ME, editors. Developmental psychology: an advanced textbook. 4th edition. Mahwah (NJ): Lawrence Erlbaum Associates; 1999. p. 125–95.
[20] American Medical Association. Mass media violence and film ratings: redressing shortcomings in the current system. Proceedings of the House of Delegates, 143rd Annual Meeting. Chicago (IL): American Medical Association; 1994. p. 78–89.
[21] Dart B, Shepard S. Parents pan film ratings system; Standards for movies, other media have grown lax, they complain. Des Moines Sunday Register June 27, 1999;A8.
[22] Goldstein P. How 'Jack' hopped away with a PG rating. Available at: www.calendarlive.com/movies/goldstein/cl-et-gold28jan28.story. Accessed Jan 30, 2003.
[23] Thompson KM, Haninger K. Violence in E-rated video games. JAMA 2001;286:591–8.
[24] Haninger K, Thompson KM. Content and ratings of teen-rated video games. JAMA 2004;291: 856–65.
[25] Yokota F, Thompson KM. Violence in G-rated animated films. JAMA 2000;283:2716–20.
[26] Thompson KM, Yokota F. Violence, sex and profanity in films: correlation of movie ratings with content. Medscape Gen Med 2004;6:3.
[27] Walsh DA, Gentile DA. A validity test of movie, television, and video-game ratings. Pediatrics 2001;107:1302–8.
[28] Screen It! entertainment reviews for parents. Available at: http://www.screenit.com. Accessed December 5, 2004.
[29] Kids-in-Mind. Available at: http://www.kids-in-mind.com. Accessed December 5, 2004.
[30] Kunkel D, Farinola WJM, Cope KM, et al. Assessing the validity of V-chip rating judgments: the labeling of high-risk programs. In: Greenberg B, editor. The alphabet soup of television program ratings. Cresskill (NJ): Hampton Press; 2001. p. 51–68.
[31] Rideout V. Parents, media and public policy: a Kaiser Family Foundation survey. Menlo Park (CA): Kaiser Family Foundation; 2004.
[32] Gentile DA, Walsh DA. A normative study of family media habits. J Appl Dev Psych 2002; 23:157–78.
[33] Cantor JR, Stutman S, Duran V. What parents want in a television rating system: results of a national survey. Available at: www.pta.org.
[34] Cantor JR. Ratings for program content: the role of research findings. Ann Am Acad Polit Soc Sci 1998;557:54–69.

[35] Gentile DA. National survey of parent media attitudes, behaviors, and opinions. Minneapolis (MN): National Institute on Media and the Family; 1996.
[36] Bushman BJ, Cantor JR. Media ratings for violence and sex: implications for policymakers and parents. Am Psych 2003;58:130–41.
[37] Bickham DS, Wright JC. Television ratings and the viewing preferences of children: a comparison of three systems. Poster presented at the Biennial Meeting of the Society for Research in Child Development. Minneapolis, April 19–22, 2001.
[38] Gentile DA, Oberg C, Sherwood NE, et al. Well-child visits in the video age: pediatricians and the American Academy of Pediatrics' Guidelines for children's media use. Pediatrics 2004;114:1235–41.
[39] Baldwin W. Testimony submitted to the Committee on Governmental Affairs, United States Senate. Available at: http://www.senate.gov/~gov_affairs/072501_baldwin.htm. Accessed July 25, 2001.
[40] Lowenstein D. Testimony submitted to the Committee on Governmental Affairs, United States Senate. Available at: http://www.senate.gov/~gov_affairs/072501_lowenstein.htm. Accessed July 25, 2001.
[41] Rosen H. Testimony submitted to the Committee on Governmental Affairs, United States Senate. Available at: http://www.senate.gov/~gov_affairs/072501_rosen.htm. Accessed July 25, 2001.
[42] Valenti J. Testimony submitted to the Committee on Governmental Affairs, United States Senate. Available at: http://www.senate.gov/~gov_affairs/072501_valenti.htm. Accessed July 25, 2001.
[43] Taylor P. Hollywood directors push for rating-system overhaul. Available at: http://www.freedomforum.org/news/2000/09/2000-09-19-06.htm. Accessed: September 22, 2000.
[44] Walsh DA, Gentile DA. Sixth annual video and computer game report card. Minneapolis, MN: National Institute on Media and the Family; 2001.
[45] Donnerstein E, Slaby RG, Eron LD. The mass media and youth aggression. In: Eron LD, Gentry P, Schlegel J, editors. Reason to hope: a psychosocial perspective on violence and youth. Washington, DC: American Psychological Association; 1994. p. 219–50.