16

Consumer testing of food products using children

R. Popper and J. J. Kroll, Peryam & Kroll Research Corporation, USA

16.1 Introduction

Nearly one-third of the world’s population is under the age of 15. In the United States alone, there are over 50 million consumers in that age group, accounting for a youth market in excess of $300 billion, with food and beverage representing as much as 60% of that market.

The size of the business opportunity has resulted in a highly competitive environment for food manufacturers as they try to gain and maintain their share of the youth market. While parents serve an important gatekeeper function in determining which products are purchased, the ‘pester’ power and influence of even young children on food purchase decisions continues to rise. Today’s children have more choices and more influence over their parents’ purchase decisions than ever before. Children are also making many more of their own purchase decisions, with ever greater sums of money under their direct disposal. Therefore, no manufacturer can succeed in the youth marketplace today without optimizing products for their consumer target – the child. By successfully appealing to children, marketers also stand a chance of building long-term brand loyalty, as early exposure to products and brands may form the basis for brand choices later in life.

Successfully developing and optimizing foods and beverages for children requires involving children in the product development process. Children’s needs and wants differ from those of adults, and product development must be guided by insights into what uniquely motivates and appeals to them. For new products, input from children may be needed at every stage of product
development, from early idea exploration and prototype screening on through product and package optimization, as well as advertising copy development. For products undergoing reformulation, either for the purpose of cost reduction or product improvement, research may be needed to confirm that the product change is not detrimental to acceptability or does indeed improve the product.

In conducting research with children, it is important that the methods employed take into consideration the physical, emotional, and cognitive development of the children being asked to participate in the study. The assumption that methods appropriate for adults can be used with little or no modification in a study involving children is almost certain to lead to disappointing results. Furthermore, it is important to consider the specific age of the child in deciding what methods to employ, since motor skills, language skills, and reasoning abilities develop rapidly during childhood.

The primary focus of this chapter is on methods appropriate for quantitative testing of product acceptability with children who are between the ages of 6 and 12. Methods for testing pre-school-age children are also briefly considered. Infants and toddlers require methods grounded in behavioral observation, which are outside the scope of this chapter. Children who are above the age of 12 are capable of using many of the same research tools as adults. The chapter begins with a review of what is known about children’s sensory perception and how food preferences develop, areas that can have important implications for product design as well as testing methodology. The chapter then reviews the techniques for quantitative consumer testing and provides examples of the types of questions and scales appropriate for testing with children of different ages.

16.2 Sensory perception: sensitivity and perceived intensity

Are children more or less sensitive than adults to taste or olfactory stimulation? Studies of taste thresholds present a confusing picture, with some studies suggesting that children as young as 5–7 years of age have similar detection thresholds as adults, others finding that children of this age have poorer sensitivity than adults (Guinard, 2001). Various methodological differences among the studies have been suggested as the source of these inconsistencies (James et al., 1997). In their study, James et al. tested 8–9-year-old children and young adults of both sexes under the same experimental conditions, determining detection thresholds for sucrose, sodium chloride, citric acid, and caffeine. A two-alternative forced-choice procedure was used, in which respondents had to indicate which sample ‘tasted stronger’ (one of the samples contained only water, the other the taste stimulus). The fact that the thresholds did not differ across two replications was indicative of the reliability of the procedure. Girls had detection thresholds similar to the adults; boys, on the other hand, proved to be somewhat less sensitive than adults, especially in the case of citric acid. De Sousa Coelho et al. (2005) also found that boys were somewhat less sensitive to the presence of citric acid in water than girls.
Greater differences in sensitivity between children and adults are evident as the task increases in complexity. Oram et al. (2001) compared 8–9-year-old children to adults in terms of their ability to recognize a particular taste (e.g. sweet) in a mixture of two tastes (e.g. sweet and sour). While children were able to correctly identify a taste as sweet, sour, or salty when it was the only taste present, they performed markedly poorer than adults in correctly identifying the components of a binary taste mixture. Children tended to recognize only one of the components, the adults recognized both. These differences could be due to differences in taste perception, in cognitive ability (e.g. ability to separately process two sensations), or in response strategy (e.g. children may have focused on the more intense or more appealing taste quality).

Children and young adults tend not to differ much with respect to olfactory thresholds (Lehrner et al., 1999). However, while children’s olfactory sensitivity is very similar to that of adults, children are less likely than young adults to correctly identify (name) an odor or to recognize it as one that had been presented earlier in the experiment (Lehrner et al., 1999; Lehrner and Walla, 2002). The fact that children perform less well at identifying odors than young adults is not surprising. Children also perform less well than young adults in picture identification (Cain et al., 1995). Children have the olfactory sensitivity but lack odor-specific knowledge, which accumulates slowly over time. Odors may be unfamiliar to children; and, even if familiar, may not have been become strongly associated with verbal descriptors. Semantic encoding (that is, the association of smells with words) is a key component in odor memory (Rabin and Cain, 1984), especially in children (Lehrner et al., 1999; Lumeng et al., 2005). Cain et al. showed that when given the opportunity to learn odor names in a paired association task, children improve quickly in performance, although the learning curve is much attenuated when the odors are novel to begin with.

Another measure of sensory sensitivity, separate from detection threshold, is supra-threshold intensity perception. Compared with adults, children appear to perform less well at ranking the sweetness of beverages varying in amount of sucrose (Kimmel et al., 1994; De Graaf and Zandstra, 1999; Liem et al., 2004a). The rate at which perceived intensity changes with stimulus concentration may also differ between children and adults. Zandstra and De Graaf (1998) varied the concentration of sucrose in orange drink and found that for children aged 6–12 sweetness increased less rapidly as a function of sucrose than it did for adolescents or adults. Using a category scaling procedure, the children rated the low sucrose concentrations as sweeter than did the adults, but rated the high sucrose concentrations as less sweet. In a second study, De Graaf and Zandstra (1999) replicated these results for orange drink, but found that in water the sweetness functions for adults and children were similar (as was also reported by James et al., 2003). Thus, food context may affect children and adults differently. This conclusion is supported by James et al. (1999), who used the method of magnitude estimation to scale sweetness, and by Temple et al. (2002), who used a computer-based time-intensity scaling procedure. Both studies found that sweetness increased in a similar fashion for both adults and children in response
to increases of sucrose in water, but increased more slowly in orange drink and some other food applications. In the case of these more complex stimuli, children may have found it difficult to attend exclusively to sweetness and may have been influenced in their sweetness ratings by other sensory characteristics.

While sweetness is the sensory characteristic most frequently studied with children, a number of studies have been reported on supra-threshold sourness perception. These studies indicate that adults and children rate sourness similarly in response to variations in citric acid, both in orange drink (Zandstra and De Graaf, 1998) and in gelatin dessert (Liem and Mannella, 2003).

Much remains unknown with regard to differences between adults and children in sensory perception, especially in modalities other than olfaction and taste (for the few examples of research on texture perception in children, see Szczesniak, 1972; Oram, 1998; Narain, 2005). Adults and children show many similarities in sensory sensitivity, both in terms of detection thresholds and super-threshold intensity perception. In case where differences have been reported, it is important to consider the possible sources for the findings. Any determination of threshold sensitivity or supra-threshold perception in children is complicated by the fact that measuring sensory perception in children is difficult – differences between adults and children in sensory perception may reflect, in part at least, differences in how children interpret the question they are asked and how they use the scales on which the research is based. Liem et al. (2004a) provide an example of how important the question is to the experimental outcome. In their study, 4-year-old children were not able to reliably discriminate the intensity of beverages differing in sweetness, using either a two-alternative forced choice or a ranking procedure. However, when asked to state their preference, the children consistently preferred the sweeter of two formulations in a pair and were able to rank order beverages from least to most preferred. Obviously, the children could indeed discriminate among the sweetness levels, but did not understand the intensity scaling instructions.

A number of the studies on intensity scaling cited above included control conditions in which children and adults rate simple visual stimuli for appearance characteristics (such as darkness or length) using the scaling procedures used to rate taste intensity. Children and adults usually perform identically on scaling such visual characteristics, suggesting that performance differences in perception of taste intensity are not solely a result of the scaling methodology. Whether the observed differences in taste sensitivity reflect maturational differences in the sensory systems of children and adults or developmental differences in attention and cognition has yet to be fully sorted out. The fact that differences between children and adults are more likely to reveal themselves with complex rather than simple taste stimuli (e.g. with taste mixtures and real foods rather than simple aqueous solutions) suggests that higher mental processing plays at least some part in accounting for the age difference in performance.

One practical implication of these findings is that to the extent to which children and adults differ in their sensory perception, children may not notice changes in product formulation to the same degree as adults. Ingredient
substitutions and reformulations such as a reduction in sodium or a removal of trans fatty acids need to be tested with children in order to determine whether children notice a difference. Of course, whether a difference in sensory perception is of a magnitude to result in a difference in food acceptability is a separate, and usually more critical, question.

16.3 The origin of food preferences

Humans are born with an innate liking for sweet and an aversion to bitter, as has been shown by studies of reflexive facial expressions and food intake in newborns (see Birch, 1999, for a review). These genetic predispositions make evolutionary sense, since sweet foods (e.g. certain fruits) tend to be nutritious and high in energy, whereas bitter foods can be poisonous. Sour tastes are also rejected by newborns. A genetic predisposition towards liking of salt has not been so clearly established; newborns are indifferent to salt, but infants at 4 months show a liking for moderate levels of salt, possibly the result of a natural maturation process (Beauchamp et al., 1986).

Humans are also genetically predisposed to avoid unfamiliar foods (Birch, 1999). In human evolution, this food neophobia served a protective function, since unfamiliar foods could cause illness or death. However, an infant’s early experience with foods begins to counteract this neophobic response. While basic tastes such as sweetness and bitterness are intrinsically pleasant and unpleasant, the preferences for specific foods are largely learned, and the diversity of world cuisines attests to the role of culture and environment in shaping food preferences and overcoming the neophobic response (Rozin, 1984).

Many food preferences are learned based on the positive physiological consequences that follow consumption, such as feeling of satiety (Birch et al., 1999). However preferences are acquired even in the absence of such positive physiological reinforcement, based on repeated exposure to a food alone. Zajonc (1968) was the first to identify the ‘mere exposure effect’ across a variety of stimulus domains, demonstrating that simply the repeated exposure to a stimulus, such as a sound or shape, can enhance liking. The role of early experience in food preferences has been demonstrated in a variety of studies. What the mother eats during pregnancy and lactation can affect an infant’s flavor preferences (Birch, 1999), because of flavor cues contained in amniotic fluid or breast milk. Beauchamp and Moran (1984) showed that infants who were routinely fed sweetened water during the first months of life showed a greater preference for sweetened water at 2 years of age. Even bitter-tasting foods are subject to early learning effects. Protein hydrosolate infant formula (recommended for infants that do not tolerate cow’s milk) tastes bitter as well as sour and is not well accepted by infants. Mennella and Beauchamp (2005) showed that exposure to these formulas starting shortly after birth leads to greater acceptance (as measured by intake) in infants aged 5–11 months. Early experience with protein hydrosolate formulas also has consequences for food
preferences later during childhood. According to Liem and Menella (2002), children who experienced protein hydrosoluate formulas early in life showed a preference at age 4–7 for higher levels of citric acid in juice (i.e. preferred a more sour taste) than did infants with no such experience.

The effect of exposure seems to be proportional to the amount of repetition. In a study conducted by Birch and Marlin (1982), 2-year-old children were exposed to initially unfamiliar foods with varying frequency (from zero to 20 times) over a period of several weeks. Preference for foods at the end of the study was almost perfectly correlated with exposure frequency: the more frequently a child was exposed to the food, the more the child liked it.

Food preference is also subject to social influences. Birch (1980) showed that 4-year-old children were influenced in their preferences by their peers who sat next to them during lunch. For example, children that initially did not like a vegetable grew to like it after repeatedly observing a peer consume that vegetable. This change in preference was relatively long lasting, persisting for several weeks and in the absence of the peers. In general (see Birch, 1999), older children are more effective role models than younger ones, mothers are more effective than strangers, and, for older pre-schoolers, adult heroes are more effective than ordinary adults.

There are several implications — for marketers, product developers, and researchers alike — of the way food preferences are acquired. Children’s neophobia is likely to affect their willingness and response to novel foods in a research setting. Loewen and Pliner (2000) developed a questionnaire for assessing individual differences in neophobia in children that might be a useful attitudinal measurement for helping to explain children’s response to novel foods in test situations.

Marketers and product developers need to balance children’s desire for novelty with their propensity to prefer the familiar. The success of green colored ketchup in the United States, which combines a familiar flavor with an unfamiliar color, is a good example of this principle. In a laboratory context, Pliner and Stallberg-White (2000) demonstrated a similar phenomenon, showing that 10–12-year-old children were far more willing to try an unfamiliar chip when it was combined with a familiar dip than when it was presented with an unfamiliar dip.

The effect of repeated exposure on food acceptance among children suggests that product sampling and other strategies that encourage repeat consumption may help build acceptance of a new product. The learning effect also has implications for how to test novel foods with children. The limited exposure typical in taste tests may lead product development to underestimate the potential for unfamiliar flavors or textures. Finally, the demonstrated effects of social role models on food acceptance provides marketers with a continued reason to attempt to leverage peer influence through advertising or grass-roots marketing campaigns.
16.4 Difference between children and adults in food preferences

Perhaps the most widely documented difference in preference between children and adults is that children prefer a greater intensity of sweetness than adults. The heightened preference for sweetness among children was first reported by Desor et al. (1975) and was subsequently confirmed by Desor and Beauchamp (1986) in a longitudinal study, in which the same respondents were tested at two points in time – at age 11–15 and ten years later. The heightened preference for sweetness, which was reflected in the percentage of respondents who chose the sweetest of four sucrose concentration, decreased from the first test to the second, demonstrating that this preference declines with age. Using ranking or scaling methods instead of choice procedures, a number of other studies have also found that the optimal level of sweetness is higher for children than adults (Zandstra and De Graaf, 1998; De Graaf and Zandstra, 1999; Liem et al., 2004a). Children also prefer higher levels of salt than adults, as has been demonstrated both with 9–11-year-old children (Desor et al., 1975) as well as pre-schoolers (Beauchamp and Cowart, 1990). The reasons for this heightened preference for sweet or salty are not fully understood, but are believed to be rooted in development rather than the result of environmental influences.

Recent work by Liem and colleagues has revealed a segment of children with a preference for extreme sour tastes. Across a number of studies (Liem and Mennella, 2003; Liem et al., 2004b), the researchers found that about one-third of children aged 5–12 showed a preference for an extreme sour taste, in contrast to adults as well as other children who preferred lower levels of sourness. All children, regardless of their preference, were equally able to discriminate among different levels of sourness, and showed the same level of discrimination as adults. Thus, the preference for extreme sour cannot be attributed to differences in sensory perception. The ‘sour-loving’ children did differ from the other children with respect to a number of other factors: they were less neophobic, had a greater preference for bright colors, tended to experience a wider variety of fruits, and tended to like other sour foods (sour candies, lemons), behaviors that may be examples of ‘thrill seeking’ among these children.

Whether the preference for extreme sour extends to preferences for other extreme tastes has not been investigated so far, and no longitudinal studies have yet been conducted to determine how this sour preference changes with age. It is also unclear whether the preference for extreme sour is the result or the cause of the observed differences in food habits.

Adults and children differ not only in their preferred intensity for certain basic tastes. Moncrieff (1966) studied olfactory preferences among adults and children and concluded that children aged 10–14 prefer fruity over floral smells, whereas the opposite was true for adults. With respect to food texture, children seem to prefer simple, smooth textures (Urbick, 2002) and, in bread products, dislike crunchy or chewy textures, especially at a young age where chewing efficiency is lower (Narain, 2005). The preference for simple textures may be an
example of children’s preference for foods with low ‘complexity’ (Ringel, 2005).

Children and adults may also differ in the relative importance they place on the appearance, taste and texture in assessing overall acceptability. Moskowitz (1994) cites a case study on ice cream in which children were found to place the same importance on appearance, flavor and texture, contrary to adults, who placed more emphasis on flavor and texture than appearance. Tuorila-Ollikainen et al. (1984) found that children put more emphasis on sweetness in soft drinks than on any other attribute, consistent with the demonstrated liking for sweet taste among children. Comparing children of different ages with respect to liking of meats, Rose et al. (2004a,b) found that taste and smell were of predominant importance to older children (aged 10–11), whereas texture and mouthfeel characteristics were more likely to influence acceptability in the younger children (aged 6–7), as Chambers and Bowers (1993) showed was the case for adults.

In summary, the results of published research (as well as those of unpublished industrial studies) demonstrate the importance optimizing products for children based on their distinct preferences. A product that is optimal for children is likely to differ from one that is optimal for adults – and may also differ from one that adults think would be optimal for children (Moskowitz, 1994).

16.5 Research methods for testing children

The research task and measurement technique employed in a study involving children must be age appropriate, since language, cognitive, and motor skills vary significantly by age. The Swiss psychologist Jean Piaget is well known for his description of the stages of a child’s cognitive and linguistic development. For example, Piaget distinguishes the ‘pre-operational’ stage (aged 2–6) from the ‘concrete-operational’ stage (aged 7–12). In the pre-operational stage, children are more likely to focus on a single aspect of a stimulus, whereas concrete-operational children have the ability to perceive stimuli multidimensionally.

Gollick (2002) describes some of the limitations of children that may affect their ability to answer questions in a research study. Very young children, for example, have difficulty with concept formation (e.g. sweetness) and classification (e.g. like/dislike). Even when they understand the principles, their attention span may limit their ability to perform the task. For example, 3½-year-old children may understand a sorting task, but only about half the children may have the attention span to remember the assignment and therefore successfully complete the task.

‘Seriation,’ the ability to rank things in order of magnitude, is not fully mastered until age 7, according to Gollick, and this has implications for the reliability of any scaling results from younger children. In addition, children have limited memory skills, which may affect their ability to remember a succession of flavors presented for evaluation in a sensory test. Gollick also notes the difficulty that children 6 and under have in attending to more than one
aspect of a stimulus at one time, as Piaget’s theory suggests. As a consequence, young children may attend to only one dimension of a food, unlike older children, who may be able to base their reaction on a simultaneous consideration of multiple aspects.

Unfortunately, the child’s age is far from being a perfect predictor of a child’s ability to participate in research. There is tremendous variation in skills among children of the same age. Gollick’s experience with cognitive testing has shown that the age at which 10% of children can master a particular task, compared with the age at which 90% of children can do so, varies by as much as 4 years. Thus, assumptions regarding what a particular age group can do are often going to be true only approximately, and researchers need to take into account the considerable variation in children’s abilities, even at similar ages. As Chambers (2005) has pointed out, particular caution is needed with respect to children in the ‘cusp’ years, i.e. those that are transitioning from one developmental stage to another. For example, children age 6–7 are at the border of the pre-operational and the concrete-operational stages of development, and this age group is likely to be quite variable with respect to their cognitive abilities.

Guinard (2001) has summarized published studies with regard to children’s abilities to perform a variety of sensory testing methods at different ages (see also ASTM, 2003). As one would expect, the younger the age group, the more challenging it is to devise valid, reliable test methods. These reviews conclude that children 2-3 years old are capable of expressing preference between two choices, but not much else. Children aged 4–5 are, in addition, are capable of performing attribute-based paired comparison tasks (`which is sweeter?'), ranking products in terms of preference, and rating products on simple hedonic scales. Children aged 6–7 are capable of more complex scaling tasks (e.g. intensity) and performing certain discrimination tests (e.g. triangle tests). By age 8, children are capable of performing virtually any kind of standard sensory test. By this age, children are also capable of self-administering many tests with only occasional assistance from the interviewer or experimenter. At younger ages, one-on-one interviews are usually required.

When products are expected to appeal to a wide age range, it is often convenient to test older children (8 and above), who are subject to fewer limitations regarding the appropriate research technique and who can self-administer the tests (which is less costly and time consuming than one-on-one interviewing). However, when the core target age for the product is specifically younger children, it may not be appropriate to focus exclusively on the older age group.

16.6 Hedonic testing with children

16.6.1 Hedonic scales

Being able to determine the level of a child’s liking for a product has obvious importance to product development. A number of hedonic scales for children have been proposed, some using pictures (often faces), some using words, and
some a combination of pictures and words (see ASTM, 2003). Three examples of pictorial scales, two of which are also verbally anchored, are shown in Fig. 16.1.

Kroll (1990) introduced a verbal liking scale for testing children (see Table 16.1), which has become known as the Peryam & Kroll (P&K) or the super good/super bad scale. The scale is similar to the traditional 9-point hedonic scale, except that the verbal anchors associated with the scale are more child-friendly—instead of using terms such as ‘like extremely’ and ‘dislike extremely,’ for example, it employs the terms ‘super good’ and ‘super bad.’ Testing children in the range of 5–10 years old, Kroll compared several scale variations, including the traditional 9-point scale, the 9-point P&K scale, and a 9-point face scale (similar to the one shown at the bottom of Fig. 16.1). Kroll also tested 7-point versions of these same three scales. Scales were compared on the basis of their ability to discriminate between two beverages (which paired preference

![Fig. 16.1 Examples of pictorial hedonic scales for children.](image)

Table 16.1  The traditional adult hedonic scale and the P&K hedonic scale for children

<table>
<thead>
<tr>
<th>Traditional adult hedonic scale</th>
<th>P&amp;K hedonic scale for children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like extremely</td>
<td>Super good</td>
</tr>
<tr>
<td>Like very much</td>
<td>Really good</td>
</tr>
<tr>
<td>Like moderately</td>
<td>Good</td>
</tr>
<tr>
<td>Like slightly</td>
<td>Just a little good</td>
</tr>
<tr>
<td>Neither like nor dislike</td>
<td>Maybe good or maybe bad</td>
</tr>
<tr>
<td>Dislike slightly</td>
<td>Just a little bad</td>
</tr>
<tr>
<td>Dislike moderately</td>
<td>Bad</td>
</tr>
<tr>
<td>Dislike very much</td>
<td>Really bad</td>
</tr>
<tr>
<td>Dislike extremely</td>
<td>Super bad</td>
</tr>
</tbody>
</table>
tests showed were differentially preferred). All scales found a significant difference in liking between the two beverages, but with the P&K scale the difference was more highly significant than for the face scale and the traditional scale. Across scale types, the 9-point scales discriminated better than the 7-point scales, even among 5–7-year-old children.

These results challenged two prevailing assumptions, namely that face scales were superior to verbal scales when testing children, and that shorter scales were better than longer scales. Spaeth et al. (1992), working with 8–10-year-old children, confirmed Kroll’s conclusions in a study comparing 3-point, 5-point, and 9-point versions of three scale types: the traditional hedonic scale, a face scale without verbal anchors, and a box scale verbally anchored only at the end-points. The authors concluded that children do not use face scales better than purely verbally anchored scales and do not use short scales better than longer ones. Also, anchoring the 9-point scale only at the end-points yields the same results as the traditional hedonic scale.

Face scales may actually be detrimental by introducing unintended bias or confusing the child. A face intended to show a degree of ‘dislike’ can be interpreted by a child as conveying anger, a face intended to show ‘liking’ may suggest ‘happiness.’ Children may choose the ‘happy’ face because they like it better, rather than because it represents their opinion about the food they taste. Cooper (2002) has found that the eyes and the mouth are particularly important to the interpretation of the facial expression and are more likely than other elements to lead to misinterpretations of the scale unless carefully chosen. She also indicates that there are cultural differences with respect to the interpretation of facial expression. Certain expressions are appropriate in some cultures, but not in others.

While children 8–10-years-old can effectively use verbal 9-point scales, their ratings, compared with adults, are often higher. Figure 16.2 compares the distribution of ratings for adults and children rating the same products, with children using the 9-point P&K scale, and adults using the standard hedonic scale. The children most frequently responded with the top most category of the scale, ‘super good,’ unlike the adults, who responded less positively. Other studies show that younger children will give somewhat higher ratings than older children using the same super good/super bad scale.

There are several factors that affect children’s use of scales. Scale length is an obvious one. With fewer response choices (e.g. a 5-point hedonic scale), responses are more likely to be crowded in the upper end of the scale range (Crawford et al., 2003). Crawford et al. (2005) investigated other scale factors which may affect scale usage among 8–14-year-olds. Using a 7-point hedonic scale, the authors concluded that a vertical scale orientation leads to more positive responses than a horizontal orientation, and that a horizontal scale with the positive end on the left leads to higher ratings than one with the negative end on the left. Similar scale effects have been reported for adults (Friedman and Amoo, 1999; Sauerhoff et al., 2005). These findings suggest caution is needed when comparing findings across studies that have used ostensibly identical, but differently formatted hedonic scales.
The reason that children favor the positive end of the scale more than adults is not clear. Perhaps, unlike adults, children do not feel a need to ‘hedge their bets,’ by reserving the scale end-points for future, yet-to-be tasted products. Perhaps children lack the frame of reference that adults have developed over the years, or perhaps children are simply easier to please. Whatever the reason, the tendency for children to favor the positive end of any hedonic scale may result in a ‘ceiling effect,’ especially with highly liked foods. When such ceiling effects are

![Fig. 16.2](image-url)
a concern, some researchers have found that a preference question may provide better discrimination than scaled liking. In general, however, children’s hedonic ratings discriminate very effectively among products. A recent unpublished review by Peryam and Kroll of 14 studies involving five product categories (including such popular categories as cookies and pasta), with children and adults rating the same products, found that children’s overall liking ratings (using the super good/super bad scale) were as likely to show significant differences as were the adults’ ratings (albeit not always favoring the same products). In fact, in the six cases where significant differences were found with one group and not the other, it was the children that showed significant differences, not the adults.

16.6.2 Hedonic scale structure
One potential concern with any verbally anchored hedonic scale is how the choice of verbal scale anchors affects the psychological spacing between the scale points. The development of the standard 9-point hedonic scale was supported by extensive psychometric research (Peryam & Girardot, 1952; Jones et al., 1955), which provided the basis for the selection of the phrases used to anchor the nine scale points. On the basis of an analysis of the ratings of a different words or phrases, these early investigators were able to select phrases for anchoring the nine scale points that were approximately equally spaced psychologically. As Lawless and Heymann (1998) point out, the equal interval property is important, since the analysis of hedonic scale data almost always involves the assignment of numerical values to the responses and the application of parametric statistics, which assume equal interval spacing. Crawford et al. (2005) used a semantic analysis approach with children similar to the one Jones et al. used with adults, in order to investigate the psychological spacing of hedonic phrases, including six of the phrases from the standard adult scale. While the authors do not compare children’s perceptions to those of adults, the agreement between the children’s perceptual scale values and those reported by Jones et al. for the corresponding six phrases is very high (linear correlation \( r = 0.98 \)), indicating similar semantic spacing of those phrases by children and adults.

Studies such as Jones et al. base their conclusion about the equal interval properties of the scale on studies of word perception. These studies leave open the question of whether the equal interval nature of the scale is preserved when respondents actually use the scale to rate stimuli (such as foods).

To investigate this question and to compare scale usage by children and adults, an analysis was conducted for the purposes of this chapter using a modified correspondence analysis approach. The data comprised ratings of 15 crackers, collected over the course of three test sessions. Overall liking ratings were collected from adults (\( N = 200 \)) using the standard nine-point liking scale. Children (aged 9–12, \( N = 200 \)) rated overall liking using the 9-point P&K (super good/super bad) scale. In addition to rating overall liking, both adults and
children rated the products on a number of other liking dimensions (e.g. appearance, flavor) and rated several sensory attributes using intensity scales.

Multivariate mapping techniques are often used in sensory analysis and consumer research to map products spanned by ‘attributes’ (hedonic or intensity scales). In the present case, the perceptual mapping techniques are used to learn about differences between children and adults in scale usage, in particular in their use of an overall liking scale.

Greenacre (1984, pp. 169–184) describes how correspondence analysis can be used to analyze rating data, as opposed to frequencies (the typical domain of correspondence analysis). A related technique, dual scaling, has been described by Nishisato (1980). Only recently has this approach been applied to sensory and market research data (Abdi and Valentin, 2007; Chrea et al., 2004, 2005a; Torres and van de Velden, 2007). The technique is well suited for the present purpose, because it allows for the determination of the perceptual distance between scale points.

The correspondence analysis included all rating scale results (liking and intensity), though only the results for overall liking are reported here. Figure 16.3a shows the product map, in which products that are perceived similarly are positioned close together and products that are perceived very differently are far apart. The map shows product projections for the children, the adults, and for the consensus between the two, which represents a weighted average of the ratings of adults and children. The consensus projections of the scale points for overall liking are also shown.

Children and adults appear to rate the liking of products similarly in some cases, but not others (the interpretation of the map in terms of product attributes is beyond the scope of this chapter). Adults rated all products somewhat lower than the children (in the map, liking increases from left to right and the adult means are positioned to the left of those for the children), consistent with the age effect on liking ratings discussed above. In addition to this systematic shift, there are some differences between adults and children in the relative liking of certain products (such as M, N, and G). For example, product N is the best liked product for children, but not for adults.

Figure 16.3b shows the separate projections of the overall liking scale points for adults and children. The scale points for the standard adult liking scale and the P&K super good/super bad scale project very similarly, and are spaced at approximately equal intervals, except for some slight compression in the middle of the scale. These results provide strong evidence that the two overall liking scales are highly similar measurement instruments and are both approximately equal interval scales.

16.6.3 Hedonic testing with pre-school-age children
The hedonic scales described so far are appropriate for testing children 5–12 years of age. Within that range, younger children (aged 5–7) will require assistance in completing the test, if for no reason other than that their reading ability is
Fig. 16.3  Hedonic scale structure for children and adults as determined by correspondence analysis. (a) Products (letters) are projected in the map according to the adults’ (upper case) or children’s (lower case) ratings of overall liking and other hedonic and sensory attributes (only overall liking is shown). The consensus of the children and the adults (bold capital letters) is located between each pair of projections. The numbers represent the consensus projections of the nine scale points on the overall liking scale. (b) Separate projections of the overall liking scale points for adults (dashed line, traditional nine point hedonic scale) and children (solid line, P&K hedonic scale for children), indicating similar scale structure for both age groups.
limited. As an alternative to one-on-one interviews (which are costly and time-consuming to conduct), it is sometimes possible to test 5–7-year-old children using a classroom-style group administration, provided the questionnaire is short and the test moderator guides the children through the questionnaire one question at a time.

But what methods are suitable for determining liking among children aged 3–5? Birch (1979, 1980) has successfully used a ‘ranking by elimination’ procedure, a variation of the traditional ranking procedure. According to this procedure, children first taste a number of samples and then choose their favorite. This sample is then set aside, and children re-taste the remaining samples, once again indicating their favorite. This process continues until all samples have been ranked. While used extensively in academic research by Birch and others, the use of this method in industrial applications appears to be rare.

Kimmel et al. (1994) concluded that 4–5-year-olds were able to use a 7-point face scale. Chen et al. (1996) found that 3-year-old children were able to use a 3-point scale, 4-year-olds a 5-point scale, and children 5 years old a 7-point scale. On the other hand, Léon et al. (1999) found low repeatability among 4–5-year-olds using three different methods, including a simple binary classification (like and dislike).

Recently, Popper et al. (2002) compared two different methods for measuring liking among pre-schoolers, the ranking by elimination procedure and a 5-point bifurcated scale. In the latter procedure, the child was first asked if the sample was ‘good’ or ‘bad,’ and, depending on the answer, was then asked whether the sample was ‘really good’ (or ‘really bad’) or ‘just a little good’ (or ‘just a little bad’). If the child had trouble committing to whether the sample was good or bad, the answer was recorded as ‘neither.’

Pre-school-age children are pre-literate and must by necessity be interviewed one-on-one. Typically, research personnel (usually female) serve as interviewers. Popper et al. included two interviewer conditions: in one condition, the child’s mother was the interviewer, in the other condition, the interview was conducted by a female researcher (P&K staff person) unfamiliar to the child. When interviewed by the researcher, Mom was not present in the room with her child.

Both the ranking procedure and the bifurcated scale resulted in significant differences among the samples tested. Greater discrimination, using either procedure, was obtained when the child was interviewed by the mother than by the unfamiliar researcher (see Fig. 16.4). When Mom did the interviewing, the average ratings for the three formulations spanned a larger range than when the researcher did the interviewing, and the differences were more likely to be statistically significant. This effect varied by age – the benefit of Mom as interviewer was evident at ages 3 and 4, but was largely absent by age 5, where Mom and the P&K interviewer gave very similar results. The authors caution that using the mother as interviewer may not be preferable in all situations, especially when there is a risk that the mother could introduce her own biases about the products. In the study, the samples the child evaluated all looked the
same (and the mother did not taste them). In situations when the appearance of the samples might suggest differences in nutrient content or when brand information is provided as part of the test, the mother’s role as the interviewer would need to be carefully assessed.

16.7 Use of intensity and just-about-right scales

The research cited in Section 16.2 on intensity scaling with children demonstrates that children are capable of using intensity scales, although their ratings may differ from those of adults. In these studies, children typically become familiar with the scales using some practice tasks (e.g. scaling visual stimuli) prior to using them to rate foods. Also, these studies usually focus on only one or at most a few sensory attributes at a time.

Swaney-Stueve (2002) undertook the challenge of attempting to train children in descriptive analysis. Using six brands of peanut butter and a ballot including 14 or more attributes, she demonstrated that it is possible to train children as young as 9 years old to reliably discriminate among products using intensity ratings. The results of the 9-year-olds did not differ much from those obtained with children aged 13–14 and 16–18, although the results from the children and teen panels did differ somewhat from those obtained with two adult panels. In another example of the use of descriptive analysis with children, Narain (2005) showed that even 4-year-olds could be trained to use texture attributes to distinguish among breads. These studies demonstrate that children are capable, given appropriate training, to perform sensory tasks that are cognitively quite demanding.
In consumer tests with children, intensity scales are sometimes included in addition to hedonic scales. Intensity scales can help provide product developers information on how children perceive the differences among the products included in the test. By comparing their ratings with those of adults, these scales may also provide some insight into how children use terms such as ‘sweet,’ ‘crunchy,’ etc. In most consumer tests, children receive no or only minimal explanation of the sensory terms. For some attributes, such as sweetness, the assumption that children (typically aged 8–12) understand the meaning of the sensory characteristic they are being asked to scale is probably warranted. For other attributes, this assumption may be tenuous. The research on odor recognition and identification cited in Section 16.2 suggests that flavor concept formation and flavor naming is a learning process that extends into adulthood. Little research has been published that would tell product developers for which sensory attributes (and for which types of products) children are capable of generating meaningful intensity scaling results (in the absence of training), or what words to use to refer to different sensory characteristics (e.g. do children understand the meaning of ‘mouth coating’?). Examples from Peryam and Kroll’s research suggests that some attributes can indeed be scaled quite successfully, as was shown in the study on crackers described earlier. Figure 16.5 shows the correlation between children and adult ratings on three attribute intensity questions, using 9-point intensity scales. The agreement between children and adults was quite high.

Information regarding the preferred level of a sensory attribute (e.g. sweetness) is often obtained from children using just-about-right scales, following the common practice of using such scales in research with adults (see Chapter 17). The adult version of the just-about-right scale is usually a 5-point scale, in which the middle category is labeled ‘just about right’ and other scale points are labeled, for example, ‘too weak’ and ‘much too weak’ on one side and ‘too strong’ and ‘much too strong’ on the other. For use with children, many researchers shorten the scale to three points, ‘too weak’, ‘just about right,’ and ‘too strong’. Again, little research has been published on children’s use of such scales, or their benefits and limitations. In Peryam and Kroll’s experience, just-about-right scales can provide meaningful results with children, although as in the case of intensity scales, careful consideration must be given to the types of attributes children are asked to evaluate. For example, children appear to be able to use just-about-right scales to flag product issues regarding appearance, size, or visual amount of an ingredient, as well as for basic tastes and simple texture attributes. In the case of flavor (other than basic tastes) and more complex appearance and texture terms, just-about-right scales tend to be less informative with children, most likely because of the difficulty children have understanding these attributes. Scaling these attributes often results in a high percentage of ‘just right’ responses, especially among younger children (aged 6–7), compared with adults using the same 3-point scale.
Fig. 16.5  Ratings of sensory intensity of 15 crackers by adults and 9–12-year-old children.
16.8 Future trends

Given the importance to the food industry of conducting research with children there is need for more research in several areas. At the level of basic science, much still needs to be learned about children’s chemosensory abilities and perceptions. The basic tastes and olfaction have been more extensively studied than other sense modalities, such as texture, yet even in the area of basic tastes, too few studies have involved complex foods, as opposed to test solutions or simple beverages.

For applied testing, the scaling methods most appropriate for children will continue to require investigation, especially as testing technology evolves to various forms of computerized testing. The potential of leveraging technology to improve children’s comprehension and use of scales in consumer tests remains to be explored. The use of the Internet for research with children is also likely to grow. Cooper (2005) has reported on the use of a children’s Internet panel for qualitative research.

While methods for scaling hedonic scaling have been widely investigated, the benefits and limitations of using intensity and just-about-right scales with children in the context of consumer research requires more investigation, leading to guidelines for when to use such scales and how to refer to different sensory characteristics in a way easily understood by children.

Several other testing parameters are in need of further study, such as time of day. According to Gollick (2002), depending on the time of day, a child’s IQ score on a standardized test can vary by as much as 30 points. Urbick (2002) advocates conducting consumer tests with children in the morning, when kids are most alert, and avoiding after-school hours, when children are mentally tired and need unstructured playtime and a chance to be physically active. Time of day may also affect children’s reactions for other reasons. Some foods are more appropriate at some times of the day than others. Children at an early age have acquired a sense of time appropriateness for different foods (Birch et al., 1984) and the effect of this awareness on test results remains to be better understood.

Food products developed for children are unique in many respects – their formulation, their packaging, and their messaging. Sensory researchers are prone to think about the product itself, but for children the success of a new product may depend not just on its taste and texture, but on how it handles, its play value, the image it projects, and how successfully all these aspects are integrated. It is likely that food developers will need to pay increasing attention to the overall ‘product concept fit,’ which may mean introducing advertising language and brand information into product research with children more often than is done today (if it is included at all). Little research has been conducted to understand how children process such information in the context of evaluating food and how such information impacts their liking ratings (Bahn, 1989; Allison et al., 2004).
16.9 Sources of further information and advice

Birch (1999) provides a comprehensive review of research on the development of food preferences in children and the differences between children and adults in sensory sensitivity. Sensory research methods for use with children are reviewed by Popper and Kroll (2005), Guinard (2001) and Resurreccion (1998). A guide on testing with children, developed in 2003 by ASTM Committee E-18 (ASTM E-2299-03), provides a great deal of practical advice on how to conduct effective testing with children and includes recommendations regarding the age appropriateness of various sensory research methods. Journals such as Food Quality and Preference, Appetite, the Journal of Sensory Studies, and Chemical Senses are good sources for information for the latest research findings on children’s sensory perception, food preferences, and the associated research methods. The biannual Pangborn Sensory Science Symposium provides an important forum for researchers from academia and industry to share findings regarding conducting research with children.

16.10 References


KROLL B J (1990), ‘Evaluating rating scales for sensory testing with children’, Food Technol, 44(11), 78–80, 82, 84, 86.


PLINER P & STALLBERG-WHITE C (2000), ‘“Pass the ketchup, please” familiar flavors increase children’s willingness to taste novel foods’, Appetite, 34, 95–103.

POPPER R & KROLL J J (2005), ‘Conducting sensory research with children’, J Sensory
Studies, 20, 75–87.


