Chapter 10

The Truth-Value Judgment Task

Peter Gordon

From: D. McDaniel, C. McKee, H. Cairns (Eds.) *Methods for assessing children's syntax.* Cambridge, Mass: MIT Press.

10.1 The Nature and History of the Truth-Value Judgment Task

The truth-value judgment (TVJ) task has proved to be one of the most illuminating methods of assessing children's linguistic competence developed in recent years. The fundamental property of this task is that it requires the child simply to make a bipolar judgment about whether a statement accurately describes a particular situation alluded to in some context or preamble. The success of the task clearly resides in the simplicity of the child's response and the significant amount of information that can be gathered about the child's understanding of complex constructions. With this method, it is possible to evaluate the child's understanding of complex constructions, which might have seemed untestable only a few years ago.

Within this chapter I will describe two kinds of TVJ tasks: *yes/no tasks* where the child responds either "yes" or "no" to a question regarding a situation and *reward/punishment tasks* where the child either rewards a puppet for making a true statement or punishes it for making a false statement about a situation.

The yes/no type TVJ task was first seen in child language studies in a paper reported in Abrams et al. (1978) studying children's comprehension of passives and cleft sentences. In this study the authors asked children yes/no questions using passive and cleft constructions and examined the validity of their responses as a measure of their comprehension of that construction. After a relatively long hibernation, the method emerged again when Gordon & Chafetz (1986) used it to examine the acquisition of actional versus nonactional passives. More recently, Stephen Crain and several others have extensively used the reward/punishment version of the task to evaluate children's knowledge of complex syntactic principles and quantification (Crain 1991).

Of course, the psycholinguistics literature contains numerous studies that require subjects to respond either "yes" or "no" to a particular question or display such as in a lexical decision task. A key difference is that, unlike the TVJ tasks, such tasks tend to be inherently metalinguistic in nature, asking subjects to judge whether a string is a word or not, whether a word belongs to a particular category,

and so on. Other studies such as sentence verification tasks are less metalinguistic, and subjects directly answer questions about situations (e.g., Clark 1974).

A major difference in the aims of language acquisition studies and those of adult studies is that the latter tend to test for performance variables where reaction times are used to make inferences about the structure of and access to existing knowledge. In acquisition studies one is generally interested in determining whether that knowledge exists at all. Of course, there is always the possibility that the apparent lack of linguistic competence could be due to performance factors in any particular case. The TVJ task generally minimizes this possibility--if constructed properly--and may act as a demonstration that knowledge exists in the child or is perhaps structured in some transitional manner.

Other tasks used in language acquisition studies have been equally simple in terms of requiring only a bi-polar response. A classic method is the grammaticality judgment task, where the child hears a sentence and is asked whether it was "good" or "silly," or "right" or "wrong" (see Chapter 11, this volume) or where the child is told to reward a puppet who "said it right" (Gordon 1981, 1982; Hochberg 1986). Both of these methods superficially resemble a TVJ task. However, there is a fundamental difference.

In grammaticality judgment tasks and other metalinguistic tasks, there is an implicit assumption that the child understands the notion of a sentence being "good"/"right" or "silly"/"wrong" as relating to intuitions about grammaticality. The TVJ task, on the other hand, makes no such assumptions. Rather, it assumes only that the child has some conception of the notion of truth in the sense of a correspondence between what is said and the situation referred to. Crucially, this requires the child to have an idea of what was said--that is, to construct a valid interpretation of the sentence via the parsing mechanisms, grammar, pragmatic assumptions, semantic entailments, and so on, that are available to her at that point in development. The advantage is that the task does not require the child to bring any of these processes to consciousness in any explicit way.

10.2 Design Issues

Let me begin by describing the two kinds of TVJ tasks with illustrations from the literature. First, consider the yes/no version. In this task, the experimenter is attempting to obtain a measure of the child's competence with regard to a particular construction. For example, Gordon and Chafetz (1986) were interested in children's understanding of actional and nonactional passives. This interest arose from an intriguing effect originally noticed by Maratsos et al. (1979), namely, that children appeared to perform worse on nonactional passives than on actional passives. Gordon and Chafetz (1986) wanted to study whether this deficit

with nonactional passives could be due to a simple function of input frequencies for these two kinds of passives. Since children hear fewer nonactional passives, it is possible that, to some extent, they might have difficulty with nonactional passives because they initially learn passives verb by verb, rather than as a general rule. Gordon and Chafetz reasoned that if verb by verb learning was indeed taking place, then children should perform consistently over a test-retest of the same verbs.

Previous tests of the passive (de Villiers 1984; de Villiers, Phinney and Avery 1982; Maratsos et al. 1979; Maratsos et al. 1985; Sudhalter and Braine 1985) had employed either picture identification tasks or agent identification tasks. In the former task children have to choose which of two pictures correctly depicts the passive sentence being spoken by the experimenter. Clearly, depicting states such as knowing and liking is problematic. In the agent identification task the child hears a passive sentence such as "Mickey was kicked by Goofy", and is then asked, "Who did it?" or "Who kicked the other one?" Although this paradigm works well for action verbs such as *kick*, it also encounters serious problems with nonaction verbs such as *know* and *like*, where it is not clear that anyone is actually "doing" anything, since no strict agent role is associated with these verbs. In addition, there is evidence that children are less likely to remember the arguments of nonaction verbs than those of action verbs, even when these verbs are in the active voice (Lempert and Kinsbourne 1981). It could therefore be the case that differences in children's performance on actional and nonactional passives are simply an effect of memory and hence one of performance rather than competence.

The TVJ task is ideal in this situation, since it eliminates these problems. In this task the child is presented with a short story describing the events in a picture--for example, a story about a boy called John, waking up and hearing music on the radio. The child is then asked two questions like those in (1) and (2).

(1)   *Was the music heard by John?*

(2)   *Was John heard by the music?*

Let us consider the demands of this task. First, comprehension of the passive construction is revealed through the appropriate choice of argument order. This turns out to be equally natural for both action and nonaction verbs. The task involves no pragmatic anomalies associated with attempting distinguish agency within nonactional states. Second, the task requires that the child does in fact understand the passive construction in order to answer correctly.

On the other hand, the task requires no more memory load than that required for everyday discourse processing where some semantic evaluation of a sentence is required. That is, the child can evaluate the truth of the sentences

directly against the real-world situation, without having to store in memory the arguments associated with the verbs. In addition, this task allows use of nonreversible passives, where switching of arguments results in an anomaly and there is therefore no possibility that the child can get confused about who is doing what in the task. In other tasks, such as picture identification or toy-moving tasks, nonreversible passives allow the child to get the answer right by using semantic/pragmatic principles without necessarily knowing the passive construction itself (Bever 1970). Since this is not possible in the TVJ task, researchers using it can take advantage of the reduction in memory demands allowed by nonreversible constructions.

Given the design of Gordon and Chafetz's study, it was possible to examine children's consistency from one week to the next in their understanding of passives of particular verbs. The results indeed showed a high degree of verb-by-verb consistency, thus supporting the idea that the effects of actionality could be explained in terms of input frequencies and lexical learning strategies.

Next, let us consider now the reward/punishment version of the TVJ task. As an example, Crain and McKee (1985) were interested in examining children's understanding of backward anaphora in light of claims by Solan (1983) and Tavakolian (1977) that children initially show a failure to allow such processes. Backward anaphora is the process whereby a pronoun or other anaphoric element precedes its antecedent, normally a full noun phrase or referring expression (R-expression) such as *John* or *the dog*. If children do disallow such coreference simply on the basis of word order, then this fact would suggest that they employ linear rather than structural principles in formulating grammatical rules. Such a finding would suggest that at least some of children's early grammatical inductions are not structurally dependent, contradicting Chomsky (1965) and Crain and Nakayama (1987).

In Tavakolian's (1977) task, children were asked to act out a sentence like (3).

(3) *For HIM to kiss the LION would make the DUCK happy.*

The majority of 3- to 5-year-olds acted out this task by having a third animal, not mentioned in the sentence, be the referent for the pronoun. In this case, both the third animal and the duck are possible antecedents for him in the adult grammar. Therefore, the children's responses indicate only a preference for an extrasentential coreferent, not a prohibition against backward anaphora per se.

The TVJ task allows the researcher to obtain responses that are less ambiguous than those elicited by in the Tavakolian's task. This can be done by narrowing the possible interpretations and making the task as simple as possible for the child. Crain and McKee (1985) tested children as young as 2 years of age

on a task in which they saw an event occur while an experimenter described it. Kermit the Frog then made a statement about the event. Children were told that if what Kermit said was true, they were to feed him a cookie. If not, they were to feed him a rag. Clearly, the humor is a crucial element of this design since it engages the children, making both the positive and negative responses enjoyable for the child without the negative connotations normally associated with a negative response.

Using this task, Crain and McKee (1985) were able to produce unambiguous evidence that 2- to 5-year-old children were quite willing to allow backward anaphora in sentences, but only in cases where this was allowed by the structural principles of grammar (Principle C of Chomsky 1981). Thus, they rewarded Kermit for statements such as (4), where *the lion* is coreferential with, but is not C-commanded by the pronoun *he*. On the other hand, they gave Kermit the rag for sentences like (5), where *the Smurf* is coreferential with the pronoun, which C-commands *the Smurf*, violating Principle C. This ingenious study provides powerful evidence that children use structural rather than linear principles in grammar formation and that, in this case, the principle they use appears to be the correct adult form that is unlikely to have been learned through modeling or instruction (see Crain 1991).

(4) *When HE stole the chickens, the LION was inside the box*

(5) **HE ate the hamburger when the SMURF was inside the fence*

10.2.1 The Role of Context and Plausible Denial

Crain et al. (1995) have recently made strong methodological arguments about the design of the TVJ task and the kind of context that should be provided. They carried out a series of studies in response to reports that children misinterpret sentences containing universal quantifiers such as *every* (Philip 1991, 1992; Roeper and de Villiers 1991; Takahashi 1991). In one of these tasks, children are shown a picture of three farmers each feeding a donkey, plus one donkey not being fed. They are then asked the question in (6). Children typically respond "no," pointing out that one of the donkeys is not being fed by a farmer. Similarly, when the picture shows three donkeys and four farmers, they respond negatively to the question in (7).

(6)   *Is every farmer feeding a donkey?*

(7)   *Is a farmer feeding every donkey?*

One explanation of this error is what Philip (1991, 1992) calls the "symmetrical interpretation." This account states that in universally quantified sentences children require that the elements be in one-to-one correspondence

regardless of the position and scope of the quantifier.

Crain et al. (1995) point out that in Philip's procedure--and in other studies on this topic--there is a pragmatic problem with the task. That is, when children see an "extra element" in the picture, either a donkey or a farmer, there is a pragmatic bias to suspect that this extra element is there for some reason, and if children do take the existence of the extra element to be relevant they can only indicate this by responding negatively. Crain et al. suggest that although children may be grammatically competent, they may override this competence in favor of pragmatic strategies. They also point out that denial in this context can only be interpreted within the symmetrical account.

Crain et al. devised a set of stories that again contained an extra element; however, this element was generated within the context of the story, not as a ruse by the experimenter to fool the child. For example, one story went as follows:

> *A mother and her two children go skiing, and come to the ski lodge afterwards for a drink. There are cups of hot apple cider and bottles of soda set out on a table. The mother takes a cup of apple cider, but the children are tempted by the sodas. The mother persuades the children to drink the apple cider instead because it will warm them up. The children each have a cup of apple cider.* (p. 33)

Kermit the Frog then described the situation:

(8) *Every skier drank a cup of hot apple cider*

Children rewarded or punished Kermit according to whether they thought what he said was true or not.

What is crucial in this scenario is that when the children in the story took their apple cider, there were extra cups of cider left on the table. This is functionally equivalent to the extra elements in the donkey studies. What was different about these two scenarios? Crain et al. explain that the difference is that the cider story provides a so-called context of plausible denial: "...if circumstances had been different in an obvious way it would be appropriate to deny the test sentence" (p. 19). For example, if there had not been enough cider cups, then the child could deny that every skier was drinking cider.[1] Crain et al. propose that if the contexts are set up with sufficient detail to allow a plausible denial, then children will not resort to pragmatic strategies in attempting to account for the extra elements in the story.

Within the set of studies that Crain et al. describe, the notion of plausible denial is supported by the fact that the manipulation worked, and children who had previously shown a symmetrical interpretation in the donkey task now showed the correct adult interpretation. Unfortunately, it is not known whether

this change was due to the existence of plausible denial, as Crain et al. suggest, or to a host of other differences between these two tasks. For example, it could be hypothesized that the reason that children ignored the extra cups of cider in the latter study is simply that the cups were backgrounded and did not figure significantly in the story. On the other hand, when the extra donkeys or farmers are presented statically within the donkey pictures, they are foregrounded and likely to distract the child into thinking that they are relevant.

Although this explanation does not deny the child's competence in quantification as demonstrated by Crain et al.'s method, it does pose a question about the source of the superior performance on the cider task as compared to the donkey task. This is significant in the present context because it raises another question, namely, whether the notion of plausible denial should necessarily be incorporated into a prescription for the design of TVJ experiments. At present I remain agnostic on this issue in the absence of confirming evidence. For example, if a researcher using Crain et al.'s procedure were to point out the undrunk cups of cider to the child subjects, would the results be the same? Similarly, if in the donkey task the extra elements were somehow backgrounded, would children still deny the questions in (2) and (3)?

Since it is not possible to state which factor (plausible denial or backgrounding) is crucial here, the design of experiments should take both factors into account until empirical support for one position is available. If the aim of a study is to determine maximal competence in children, then, like Crain et al., the experimenter should both create situations that provide plausible denial and not include elements within a task that serve only to distract the child from a correct interpretation.

On the other hand, the fact that children are so easily distracted by irrelevant information –of the type in the donkey-farmer pictures—is interesting in itself. In choosing whether to include irrelevant or distracting elements in a task, the experimenter must be clear about what questions are being asked. If the study is investigating children's maximal competence, then extraneous elements should be minimized and backgrounded. If the focus is on processing or pragmatic differences between adults and children, then it is legitimate to include such elements, though conclusions should be drawn cautiously. In other words, having distracted children into making an invalid response, the experimenter should not then claim that children therefore lack grammatical principles, even though they may fail to show evidence of them within that particular task.

10.2.2 Choosing a Truth-Value Judgment Task

Another issue raised by Crain et al. (1995) is that Philip's (1991, 1992) method is not well suited to revealing competence in that it uses only static pictures that

cannot provide the necessary context to allow for plausible denial. By contrast, in Crain et al.'s study toys were used to act out the preamble to the TVJ task. The latter procedure not only allows a full development of the preamble, but also is very useful in holding the child's attention.

In the past, the use of static pictures has generally been associated with the yes/no task, and the use of extensive stories with props has generally been associated with the reward/punishment design. However, these tasks and contexts are not necessarily linked. A yes/no question could easily follow an extensive preamble acted out with toys and props and a true or false statement could easily be made by a puppet describing a static picture.

Of course, the experimenter should decide which format is most appropriate for the task at hand. In a particular case, for example, constructing a yes/no question out of the relevant structure might be thought to add extra processing demands on the child, and therefore it might be more prudent to use the reward/punishment method. Probably, if there are going to be any such limitations, they are more likely to show up in the yes/no task; the reward/punishment task is more adaptable in this regard. On the other hand, it is often harder for a single experimenter to carry out the reward/punishment task, since usually in this design one person manipulates the toys and tells the story and another person works the puppet. Although it is probably possible for a single experimenter to carry out this task, an experimenter faced with working alone might favor the yes/no task for simplicity's sake.

Of course, the choice between formats will also be determined by the kind of construction being tested and the relationships of various elements within the task. Thus, in a study like Gordon and Chafetz's (1986), where extraneous elements do not play a role, it is quite satisfactory and much simpler to use pictures rather than having extended story lines acted out with toys.

On a recent field trip to the Kadiweu tribe in southern Brazil I became aware of other problems that can arise in choosing a TVJ design. Filomena Sandalo and I were interested in studying the acquisition of binding in this language because it admits structures that appear to violate Principle C. Thus, a sentence such as *HE-said that JOHN washed the dishes* allows the pronominal element (a verbal affix) to be corefer with the R-expression, *John* (Sandalo 1995). We started by using a puppet with the reward/punishment design. Unfortunately, the children had never encountered puppets before and reacted with a mixture of curiosity and fear that often led to tears before we changed the design. To add further problems, we could not resort to a yes/no design because Kadiweu has no words for *Yes* and *No* and therefore does not allow yes/no questions. We ended up asking the children, "who said it?" for the above sentence, which worked more satisfactorily. For those who venture beyond familiar borders, it is advisable to

consider such factors before leaving.

10.3 Construction of the Truth-Value Judgment Task

Having outlined some of the basic issues relating to TVJ tasks, I will now discuss the actual construction of a TVJ task for any hypothetical structure.

10.3.1 Training

Training for the TVJ task, although not demanded, is often desirable. For example, there is a general tendency for adult subjects to have a positive bias in verification tasks, being faster and more accurate to respond to true rather than false statements (Wason 1961). It seems plausible that children might also have such a disposition to favor encoding the truth of a statement relative to a situation rather than its falsity. The reward/punishment paradigm fosters the likelihood that giving each type of response will be equally attractive to the child by making the negative response funny (e.g., feeding a rag to Kermit). Unfortunately, it is not clear that in such cases the negative response might not become too attractive and tip the balance. In cases of doubt, data from a training study could provide evidence about whether such biases exist in the task. Such training could also encourage children to produce a balance of positive and negative responses rather than perseverating with a single type of response. Crain and Thornton (1990) report that in order to prevent this kind of pattern, they included many trials, some for which the appropriate response was negative and others for which it was positive. However, there is some advantage to getting this pattern established prior to rather than during the main phase of the experiment.

Gordon and Chafetz (1986) included a training phase in their study of passives for reasons having to do with the design of the experiment. In that study, they required that children answer both true and false versions of the yes/no question (see (1) and (2)). This was done in an attempt to eliminate some of the random responses that children might be making because of inattention. That is, if a child is not paying attention to the passive structure and is responding somewhat randomly to a question, then he will have a 50% chance of responding correctly on any single trial. On the other hand, anticipating that there will be a second item where the structure is reversed, the child may focus on the details of the construction more carefully.

Prior to testing, children were told that they would be asked two questions about a story and picture, and that they were to respond "yes" to one and "no" to the other. In the training phase, a simple construction was tested: children were shown a picture of, say, a box on a table and were asked, "Is the box on the table?" and "Is the table on the box?" With these simple items, it was possible to get children to think in terms of responding "yes" to one question and "no" to the other. In addition, if a child failed to differentiate responses, it was possible to

correct this problem during the training phase rather than during the main test and thereby to avoid losing data.

10.3.2 Context/Preamble

In all TVJ tasks the function of the preamble is to set up a situation in which production of the construction being tested is natural and comprehensible, and violates no pragmatic constraints or felicity conditions. In addition, the preamble should be constructed in such a way that at least two viable interpretations of the test construction are possible. Although one of these may be the only correct interpretation, it is crucial that, were the child to have an incorrect grammar, the situation make available an interpretation that is consistent with that incorrect grammar, and that the appropriate response--given the incorrect grammar--be the opposite of the response arising from the correct grammar. In addition, it is desirable that the situation depicted in the preamble make each of these interpretations equally salient, independent of the actual test sentence.

Of course, this begs the question of how to equate for salience. It is sometimes impossible to do this within a single context, since it may be that the correct response is to reward the puppet for a statement that correctly describes the scene as in the backward anaphora study of Crain and McKee (1985). In order to show backward anaphora, the child must accept Kermit's statement about the situation in (4), which is perhaps more salient than rejecting it, given a bias toward positivity. In such a case, it is crucial that the positive responses be counterbalanced with equally salient negative responses, as in (5). In this case, then, it is the pattern of both positive and negative responses rather than a single response type that supports the hypothesis.

An example showing where it is possible to equate for salience within the preamble comes from a study investigating the Strong Crossover Constraint (Crain 1991). In this study, children were told the following story using toy figures:

This is a *Sesame Street* story with Big Bird, Bert, and Grover. In this story they are all going for a walk with RoboCop and this Ghostbuster just before dark. The problem was that mosquitoes came out at dark, and bit everyone except for RoboCop and the Ghostbuster because they have metal suits on. Big Bird got the most bites, and is having trouble scratching them. RoboCop and the Ghostbuster say, "We'll help you. We don't have any bites." Bert says, "I don't need RoboCop and the Ghostbuster to help me. I can reach my bites." And Grover says, "Me neither. I don't need RoboCop and the Ghostbuster to help me."

Kermit then commented:

(9) *I know who THEY scratched. BERT and GROVER.*

If children obey the Strong Crossover Constraint, they should reject this sentence. Notice that there are two equally salient antecedents for <u>they</u> in this context. On one interpretation, <u>they</u> refers to RoboCop and the Ghostbuster; this antecedent is false within the context of the story and thus should lead to rag feeding. On the other interpretation, <u>they</u> refers to Bert and Grover; however, this antecedent violates Strong Crossover and should not be entertained by children who adhere to this constraint. It is possible to determine more exactly whether children in fact do entertain only the first interpretation by asking them to correct Kermit's statement.

Notice that the preamble includes a strong context for interpretation within an incorrect grammar, namely, a context where Bert and Grover are the antecedents for <u>they</u>. Since many grammatical principles are negative (i.e., stated in the form of constraints), one is often interested in constructions that violate them. If children obey the constraints, then they will fail to make the interpretation under which the principle is violated and thus reject the sentence. One can provide strong evidence for the existence of a principle if children consistently show evidence for the interpretation that does not violate the principle. This can be achieved by asking a follow-up question to verify that the child did in fact make only one of the possible interpretations. For example, one can ask the child to correct Kermit, or one can probe the child about certain elements in the story. In fact, follow-up questions are recommended in any study, since they can provide more data on the child's knowledge. The only drawback is that the follow-up question might cause the child to think that the original response was incorrect and should therefore be changed on this or future items. Therefore, such follow-up questions should be positive in tone and seem like a natural continuation of the preceding discourse.

There is no strict formula for determining how much context should be provided for a particular task. This needs to be determined on the basis of the construction being evaluated and careful consideration of the demand characteristics of the task itself. These considerations should include the following elements:

1. There should be a plausible context in which it would be "natural" for a puppet to be using that construction.

2. The story should provide possible states of affairs that could support both positive and negative responses to the question.

3. Processing demands on the child should be minimized.

4. Caution should be exercised when including "distractor" elements in the task. These are elements that are superfluous to the correct interpretation of the sentence. A child who attends to those elements in responding might have only a

tenuous control over grammatical principles but not necessarily lack those principles altogether.

These kinds of issues are by no means unique to the design of TVJ tasks. What is perhaps unusual is that with TVJ tasks it is possible to be very careful in dealing with many of these concerns by providing the right kind of context. For example, one of the most important means of reducing processing load is to provide a very vivid display with toys and props that can act as a support for memory demands. If the context is available in front of the child and can be accessed without memory demands, then it is less likely that such performance factors will play a role in the child's response. Attentional demands are also a consideration. Making the stories as entertaining as possible, using familiar figures from current child culture, and making the response as entertaining as possible crucially maintains the child's attention and reduces the possibility that performance factors will intrude on the responses.

The same basic principles apply if the stimuli are limited to pictures. These should be bright and colorful, but not too "busy" (i.e., they should not include too many elements that are irrelevant to the story and that the child might pick up on inadvertently). Of course, if the aim is to distract the child with such elements, then, as mentioned previously, one must be careful to limit the interpretation of the results with regard to children's true competence. I might also note, on this topic, that quite subtle changes in the design of a picture can radically affect the way children construe the situation being depicted. For example, Lempert (1984) has shown that children are more likely to use passives to describe pictures where the patient is colored in and the agent is not. Presumably, use of the passive indicates some kind of perspective shift toward a foregrounding of the patient over the agent. One can imagine other foregrounding and backgrounding cues such as proximity, relative size, animacy, and potency, which could affect the way that children construe the relationships within the picture and might in turn affect their responses. If possible, such asymmetries should be avoided unless they are a manipulation of the study.

10.3.3 Problems and Pitfalls

What problems might arise in designing a TVJ task? One type of problem pervasive in psychology is the so-called Clever Hans effect, where the experimenter unwittingly cues subjects into producing the correct answer and thereby overestimates their competence. Such cues can be extremely subtle, in the form of facial expressions, gaze, tone of voice, fluency differences, and a whole host of related forms of body language. Unlike researchers working in laboratory settings, researchers working with children in natural settings do not usually have the luxury of isolating the experimenter from the subject or being blind to the conditions. Attempts to restrict contamination in this way are likely to arouse

feelings anonymity, boredom, and fear in the child and therefore lead to inattention. Such concerns appear to be no more prominent in the TVJ task than other methods of studying language acquisition. However, since the TVJ task might require the experimenter to produce an ungrammatical, inappropriate, or infelicitous statement in relation to the context, it would be well to practice making such statements in a fluent manner so they are indistinguishable from statements that do not violate any of these standards (see chapter 6, this volume). In addition, if there are several possible interpretations for a particular construction, it is advisable to keep one's eyes fixed on the child and averted from the display while waiting for the child's response. This will help avoid providing subtle cues to the child. Since students who help in conducting experiments are often relatively inexperienced, they should certainly be made aware of such possible effects and be trained to avoid telegraphing correct answers.

Underestimating competence is also a concern. Underestimation can result from several factors alluded earlier. Primarily, though, it is crucial that the preamble not be made overly complex so that the child loses track of the participants and their actions. It should include as many participants and events as are necessary to set up the target construction, but no more. In addition, the target construction should be designed to be as simple as possible while still requiring the child's knowledge of the appropriate grammatical construction.

10.3.4 Design Issues and Data Analysis

Since statistical issues are dealt with in detail in chapter 14 of this volume, here, I will simply outline the major kinds of analyses resulting from the TVJ task and suggest the appropriate kinds of tests. In general, unless the results are completely categorical (e.g., all responses were as predicted), it is advisable to produce standard statistical analyses to support one's conclusions.

In designing an experiment, it is generally advisable to prepare as many items as are practical for testing; and, if possible, the number of test and control items should be equal. In the TVJ task this might amount to preparing equal numbers of positive and negative items. The limits on how many items one should use are dictated by factors such as not wanting to bore the child with repetitious procedures or make the procedure so long that the child cannot attend well by the end of the task.

One must also consider whether the task will be between subjects or within subjects. Within-subjects designs allow more powerful statistical analyses, but are subject to possible contamination effects between conditions. If one is concerned that the response to a particular construction might affect the child's response to another construction, one might consider having separate groups of children participate in these two conditions, matched as closely as possible for

factors like age and sex. Alternatively, one could stick to a within-subjects design and order the two conditions so that children receive one type of item on the first half of the test, and the other type of item on the second half of the test with both orders being represented across subject groups. One can then test for contamination by looking at group differences for each order. If there are differences, then one might consider analyzing only the first half of the items for each child, using between-subjects analyses.

A typical TVJ task results in several data points for each child. Usually, the number of items for which the correct answers are positive and the number for which the correct answer is negative should be equal. Given this scenario, the analyses are relatively straightforward. One can first carry out a t test comparing the number of "yes" or reward responses to the target positive items compared to "yes" or reward responses to the target negative items. Alternatively, one might be comparing different kinds of items and therefore be interested in the number of correct responses across conditions. For example, in comparing action with nonaction passives, Gordon and Chafetz (1986) analyzed the number of correct responses per condition (action vs. nonaction). In this case there was no positive or negative target; rather, the pattern of "yes" and "no" responses over the two questions indicated whether the child's response was correct or not. In this case the analyses are basically the same: numbers of correct response patterns are compared across conditions using the t test.

The choice of a related or independent t test will depend on whether the study uses a within-subjects or between-subjects design, respectively. If the experimenter wishes to look at multiple factors and interactions such Condition x Age, then the relevant analyses of variance should be used (see chapter 14, this volume). If the study uses a within-subjects design, then a repeated measures analysis of variance should be used. This is sometimes found in the multivariate analysis-of-variance programs of statistical packages such as SPSS. Analyses should ideally be carried out on both subjects and items. This is to ensure that the results are generalizable not only to other subjects, but also to other sets of items of the same kind. H.H. Clark (1973) recommends computing the Min F', which is a composite of these two analyses and is somewhat more stringent. However, separate subject and item analyses are usually considered adequate for making inferences about the generalization of results.

As an example, let us look at the backward anaphora study of Crain and McKee (1985). In this design the experimenter presents the child with several sentences like (4) where backward anaphora is allowed (*When HE stole the chickens, the LION was inside the box*); such sentences are target positive items. In target negative items backward anaphora is disallowed, as in (5) (*HE ate the hamburger when the SMURF was inside the fence*).

Assuming that children only reward or punish Kermit and complete all of the items, the experimenter can use the raw number of reward responses for each condition, which is the same as considering the number of punishment responses. Since children receive both kinds of items, this is then a within-subjects design utilizing the related t test. If the numbers of target positive items and target negative items differ (which should be avoided in any event), then the proportion of reward responses to each of these conditions should be used instead of the raw scores.

Some children may produce irrelevant responses such as feeding Kermit both the cookie and the rag, or they may not respond at all. In such cases irrelevant responses should be discounted and scores calculated as a proportion of the usable responses from that child. If the number of usable responses becomes less than about 75% of the items, then the experimenter should probably consider dropping that subject from the analysis and noting in the results that the subject had been dropped. A similar principle would hold for children who leave the experiment early, but again complete at least 75% of the task. An exception might be made if the several children showed a pattern of irrelevant responses associated with one particular kind of structure. In such a case, rather than discarding these children's data, the experimenter might consider performing a test comparing the proportion of irrelevant responses for that structure compared to other structures tested in the experiment.

Giving children proper training before the main test is one way to avoid such problems--ensuring that they understand what is required in the task. Running a few pilot trials is also useful in ensuring that the task runs smoothly and that the experimenters are fluent and fluid in presenting the context associated with the task, manipulating puppets, and so on.

To avoid having subjects leave prematurely, one should not overload them with items; yet one should not underestimate their capacity, either. The number of items to be presented will depend on several factors such as the age of the children, the length of time it takes to present the context, the amount of training required, and the availability of the candidate items in the language. For example, one might be looking at a particular kind of lexical item that is simply not very frequent in the language, such as irregular plurals.

Given an average preamble a few sentences long, most 3- to 5-year-olds should be able to handle up to about 20 test items plus training items. If the preamble is longer or the task appears repetitive, it is possible to halve this number for two conditions and still obtain a satisfactory number of responses for analysis. For 2-year-olds and younger 3-year-olds, who might be less able to maintain attention, the number might be limited to about 8 to 10 items. Again, pilot testing can be very useful in determining a usable number of items. If the

number becomes very small, then one might consider using a between-subjects design so that each child subject provides a useful number of data points for the condition that he or she represents.

If it turns out that only a few items can be included in the design--say three or fewer per condition--there is a concern that the data might violate assumptions about underlying normal distributions required by the t test. Although parametric tests, which make such assumptions, do tend to be fairly robust against violations of normality, one might consider using a nonparametric test such as the Mann-Whitney for between-subjects designs or the Wilcoxon for within-subjects designs. In these tests the scores for each subject are ranked ordinally against each other, and the sums of those ranks are used as the basis for determining differences.

Besides the comparison of target positive and negative items, separate analyses should be carried out for both the target positive and target negative items to determine whether the responses were significantly different from chance. The procedure here is the standard t test against what would be expected from chance responding. Normally, since there are only two possible responses, the expected chance value would be 50% for both positive and negative responses. However, the task must be carefully examined to be sure that this is indeed the case. It is possible that there is more than one way to arrive at a particular response. In this case, if chance is considered to occur at the point of choice rather than the point of response, the expected chance value might not be 50% if there are more than two choices.

For example, suppose an experimenter is testing for adherence to Principle X, sets up a situation in which the correct interpretation of the test sentence requires coreference between an anaphoric element and animal A. The alternative interpretation, which violates Principle X, can be constructed through coreference with animal B, C, or D. For example, the sentence *The dog said that the monkey is washing himself* (with a cat and squirrel looking on) requires that only the monkey can be the coreferent of *himself*. If a child pays no attention to the context and just randomly gives one of the two available response choices--"yes"/reward or "no"/punishment--then chance occurs at the point of response and responses should be evenly divided between positive and negative. On the other hand, a child who pays attention to the context, but does not know which animal should be the coreferent, might pick any animal randomly--at the point of choice--and respond appropriately according to the interpretation in which the randomly chosen animal is the coreferent of the anaphoric element. In this case the child would have a 25% chance of responding correctly and a 75% chance of responding incorrectly.

The same caution applies in cases where a construction could be many

ways ambiguous for the child who does not have the correct adult grammar. If there are two possible incorrect interpretations and only one correct interpretation, then again the probabilities should be adjusted to 67% and 33%, respectively. Ideally, the study should be designed in such a way that these problems do not arise. However, sometimes they are unavoidable and must be taken into account.

It is also sometimes useful to look at individual differences. That is, different groups of children appear to be at different developmental stages, or to have different kinds of grammars and thus may show clusters of response patterns over different sets of items or over a series of experiments. For example, in their study of universal quantifiers, Crain et al. (1995) first identified children who gave incorrect responses on the donkey task and then selected them for further testing on follow-up tasks such as the cider task outlined earlier. This kind of selectivity is one method of using individual differences to further probe linguistic competence. In using this method, however, one should always bear in mind that it involves selecting a subgroup of children and that any conclusions should not refer to "children" universally.

Sometimes one is looking for evidence of a pattern of abilities whereby competence on one construction is associated with competence on another construction (and likewise for the lack thereof). This can be demonstrated by creating a 2 x 2 matrix where columns represent construction A (pass versus fail) and rows represent construction B (pass versus fail). Each child's pattern of responses to constructions A and B will put the child into one of the four cells of the matrix (pass-pass, pass-fail, fail-pass, fail-fail). A chi-square test will show that the two constructions are significantly associated if there are more subjects on the consistent diagonal (pass-pass or fail-fail) than would be expected by chance. It is important to remember that, in the chi-square analysis, no child can contribute more than one data point to the matrix.

10.4 Appropriate Uses for the Truth Value Judgment Task

For which populations is the TVJ task most appropriate, and for which kinds of issues is it best suited to investigate? The TVJ task is adaptable for children of almost any age, so long as they are able to either answer a yes/no question or feed a puppet according to the experimenter's instructions. Of course, this does not mean that children of any age will necessarily demonstrate competence. With very young children it may be advisable to include a pretest/training phase, on the basis of which children can be selected who are able to respond appropriately when tested using the simplest of constructions. Children as young as 2 years of age can participate in TVJ tasks (Crain and McKee 1985), and certainly by 3 years of age they should be quite comfortable with the procedure. In the other direction, there appears to be no principled limit on how old subjects can be, and it is possible that a version of the procedure might even be useful for testing

adults, perhaps incorporating reaction times into the procedure. For older children, the stories should be adapted so they are age-appropriate, although the usual cast of *Sesame Street* and comic book characters are probably good for several years.

One very promising new direction is the use of the TVJ task with language-impaired populations. Early evidence of the task's utility for these populations comes from a study by Franks and Connell (in press). They report on a procedure called the video multiple interpretation task in which children are presented with videotapes of stories and are then asked yes/no questions about the events in the videos. With this procedure, Franks and Connell successfully tested children with specific language impairment on their knowledge of complex constraints on binding. One might speculate that the TVJ task would be well suited for studying any kind of population for which the experimenter wants to minimize demands on the subject in areas such as production. Adult aphasics and language-impaired children could be tested for underlying competence that might be hidden by evaluations based on spoken language or other comprehension tasks. The task might also be adapted for second language learners whose lack of spoken competence may hide greater grammatical knowledge than is immediately evident.

The TVJ task appears to be extremely adaptable for testing many kinds of constructions, and in many cases is the best choice. In a few cases it is less appropriate however; I will consider these first. Primarily, the TVJ task tests a child's interpretation of constructions relative to that child's current grammar. It cannot actually test the child's judgment about the grammaticality of that construction, nor should a correct answer be taken to indicate such a judgment. In addition, the TVJ task cannot test for production, since the child is not required to produce anything more complex than a "yes" or "no" response, or the action of feeding a puppet. On the other hand, it is possible to follow up the task with a correction by the child. Thus, if Kermit is fed the rag, he may ask the child to tell him what he should have said, or to elaborate on elements of the story, and thereby elicit the appropriate construction under consideration.

The TVJ task is limited to testing constructions that are embedded within statements rather than questions. This may seem odd since one version of the task relies on yes/no questions. However, within the TVJ task the yes/no questions are questions about the statements embedded within those questions. Thus, *Was John heard by the music?* is a question about the truth of the underlying statement *John was heard by the music;* it is not a test of the question form itself.

Much of current linguistic theory focuses on the allowability of *wh*-movement in complex structures. To test the child's knowledge of the principles that license or prohibit such movement requires that the child respond in some

way in relation to the question qua question. Since *wh*-questions do not have truth-values, it is difficult to study them using the TVJ task. Perhaps the only way to do this is to embed the appropriate *wh*-question within a matrix statement--for example, in a report of what some other character in the story had asked about. For example, Kermit might say something like, "Grover knows who the lion kissed before eating."   In this case the story might relate that one animal ate fruit and another was eaten by the lion, that both were kissed first, and that Grover witnessed only one of the kissing episodes. Although this could work, it seems cumbersome and the construction might be more easily tested by other methods. McDaniel and McKee (1992) report on use of a task examining the Strong Crossover Constraint. In this case Kermit was asked a question and answered it, and the child rewarded or punished him based on the validity of his answer.

With these few reservations, the TVJ task is an invaluable tool for testing children's knowledge of language. It can be used to study all aspects of grammar and is particularly suited for testing syntax and sentence-based semantics. The task lends itself well to constructions that involve statements but is more cumbersome for constructions like *wh*-questions. The heart of the design of TVJ tasks lies in the context provided for the test sentence. Although the experimenter must be very careful in constructing these contexts, they can set up a test that provides crucial information concerning the child's linguistic competence in many populations. This relatively new technique is fast becoming the method of choice for testing many kinds of linguistic constructions in children. Given the task's many advantages, this trend should continue.

Note

1. Notice that this definition of plausible denial differs from the traditional notion where negative constructions are used only if the affirmative is the default assumption (Wason 1965).

References

Abrams. K.H., Chiarello, C., Cress, K., Green, S., & Ellett,N. (1978).   The relation between mother-to-child speech and word-order comprehension strategies in children. In R.N. Campbell and P.T. Smith (Eds.). *Recent Advances in the Psychology of Language, Vol. 4a.   Language Development and Mother-Child Interaction*. New York: Plenum Press.

Bever, T.G. (1970) The cognitive basis for linguistic structures. In J.R. Hayes (Ed.) *Cognition and the development of language*. New York, NY: Wiley.

Chomsky, N.A. (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: M.I.T. Press.

Chomsky, N.A. (1981). *Lectures on government and binding*. Dordrecht, Holland: Foris.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335-359.

Clark, H.H. (1974). Semantics and comprehension. In T.A. Sebeok (Ed.), *Current trends in linguistics, Volume 12: Linguistics and adjacent arts and sciences* (pp. 1291-1498). The Hague: Mouton.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioural and Brain Sciences*, **14**.

Crain, S. & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Berman, J-W Choe, & J. McDonough, (Eds.), *NELS 16* (pp.94-110) GSLA, University of Massachusetts, Amherst.

Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, **63**, 522-543.

Crain, S. & Thornton, R. (1990). *Levels of representation in child grammar*. Paper presented at the 13th GLOW Colloquium, Cambridge, England.

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., & Woodams, E. (1995). Quantification without qualification. Unpublished manuscript.

de Villiers, J.G., (1984) *Learning the passive from models: Some contradictory data*. Paper presented at the Ninth Annual Boston University Conference on Language Development, October 1984, Boston, Mass.

de Villiers, J.G., Phinney, M., & Avery, A. (1982). *Understanding passives with non-action verbs*. Paper presented at the Seventh Annual Boston University Conference on Language Development, October 1982, Boston, Mass.

Franks, S. & Connell, P. (in press). Knowledge of binding in normal and SLI children. *Journal of Child Language*.

Gordon, P. (1981). Syntactic acquisition of the count/mass distinction. *Papers and Reports in Child Language Development*, **20**, Stanford University Linguistics Department, Stanford, California.

Gordon, P. (1982). *The acquisition of syntactic categories: The case of the count/mass distinction*. Doctoral Dissertation, Department of Psychology, M.I.T.

Gordon, P. & Chafetz, J. (1986). *Lexical Learning and Generalization in the Passive Acquisition*. Paper presented at the Eleventh Annual Boston

University Conference on Language Development, October 1986.

Hochberg, J. (1986). Children's judgments of transitivity errors. *Journal of Child Language*, **13**, 317-334.

Lempert, H. (1984). Topic as Starting Point in Syntax. SRCD Monographs, 49, 1-73.

Lempert, H. & Kinsbourne, M. (1981). How young children represent sentences: Evidence from the superiority of noun recall from action as compared to stative sequences. *Journal of Psycholinguistic Research*, **10** (2), 155-166.

Maratsos, M.P., Kuczaj, S.A., II, Fox, D.E.C., & Chalkley, M. (1979). Some empirical studies in the acquisition of transformational relations: Passives, negatives, and the past tense. In W.A. Collins (Ed.). *Minnesota Symposium on Child Psychology, Vol. 12*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Maratsos, M.P., Fox, D.E.C., Becker, J.A. & Chalkley, M.A. (1985). Semantic restrictions on children's passives. *Cognition*, **19**, 167-191.

McDaniel, D., and McKee, C., (1992). Which children did they show obey strong crossover? In H. Goodluck and M. Rochemont (Eds.) *Island Constraints: Theory, Acquisition, and Processing*. Dordrecht, Holland: Kluwer, 275-294.

Philip, W. (1991). Spreading in the acquisition of universal quantifiers. In Proceedings of West Coast Conference on Formal Linguistics, 1991.

Philip, W. (1992). Distributivity and logical form in the emergence of universal quantification. In Proceedings of the second Conference on Semantic and Linguistic Theory. *Ohio State Working Papers in Linguistics*, **40**, pp. 327-346.

Roeper, T. & de Villiers, J. (1991). The emergence of bound variable structures. In *University of Massachusetts Occasional Papers: Papers in the Acquisition of WH*, GLSA, Amherst, MA.

Sandalo, F. (1995). *Kadiweu as a pronominal argument language*. Paper presented at the annual meeting of Linguistics Society of America, 1995.

Solan, L. (1983). Pronominal reference: Child language and the theory of grammar. Boston, Mass: Reidel.

Sudhalter, V. & Braine, M.D.S. (1985). How does comprehension of the passive develop? A comparison of actional and experiential verbs. *Journal of Child Language*, **12**, 455-470.

Takahashi, M. (1991). Children's interpretation of sentences containing 'every'. In

University of Massachusetts Occasional Papers: Papers in the Acquisition of WH, GLSA, Amherst, MA.

Tavakolian, S.L. (1977). *Structural principles in the acquisition of complex sentences*. University of Massachusetts, Graduate Linguistics Association.

Wason, P.C. (1961). Responses to affirmative and negative binary statements. *British Journal of Psychology*, **52**, 133-142.

Wason, P.C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, **4**, 7-11.