Interaction of particle and prosody in the interpretation of factual Dutch sentences

Robert S. Kirsner, Vincent J. van Heuven and Renée van Bezooijen

1. Introduction

Aims of this study. Lexical items, such as shoe or eat (with a fairly constant and concrete reference) are easier to define than are grammatical or partially grammaticalized elements, such as verb tenses or the 'pragmatic' or 'discourse particles' (for the terms, see Levinson 1987: 129). This latter type is well-represented in such languages as Dutch, German, and Russian, but poorly represented in English. Furthermore, while certain kinds of particles in the particle-rich languages have attracted a great deal of attention (most notably scalar particles, modal particles, and perspectivity operators), others have not (cf. Abraham 1986; Van der Auwera and Vandeweghe 1984). The present study focuses on the Dutch utterance-final hoor, until recently very much a linguistic stepchild.

As an operational strategy, we prefer a monosemous rather than a polysemous analysis of hoor (cf. Kirsner 1993: 81-82). To the extent that particles such as hoor are grammaticalized, we analyze them the way one would other indisputably grammatical elements (e.g. verb tenses or pronouns) as entering into a finite number of paradigmatic contrasts. In the present case, leaving out other candidates such as joh and zeg, hoor is directly opposed to hè, with which it never co-occurs (Kirsner and Deen 1990: 3,9).

At this juncture we must note that a crucial and as yet unexplored ingredient in the interpretation of sentences containing hoor has been intonation. In previous, exploratory work (Kirsner 1991, 1993) native consultants reported that they often found themselves unable to determine the acceptability or interpretation of a printed sentence ending in hoor unless they were allowed to assume a specific melody for the sentence. This observation suggests that the kind of abstract interactional meaning communicated by particles resembles that of intonational features. With intonation, too, the statement of stable meanings has remained elusive. Analysts postulating explicit meanings for particular Dutch intonation contours have been obliged to argue at great length that the specific melodies in question really did signal the meanings proposed (cf. Keijsper 1984). Finally, it has been claimed that the messages or interactional work achieved with particles in one language may be expressed with intonation in another (e.g. Schubiger 1965).

To have any realistic basis for analyzing hoor it is therefore imperative to elicit semantic judgments under carefully controlled conditions using spoken
rather than written sentences. In order to understand how and why particle and prosodic pattern interact in the interpretation of spoken sentences, we shall postulate abstract meanings for *hoor* as well as for two specific intonational features, and from there predict and verify their separate and combined effects in terms of semantic dimensions by asking native Dutch speakers to perform a specific experimental task. This is the purpose of the present study.

Because of the highly abstract meanings that we postulate for particles and intonation features, special techniques are required to check with naive language users whether our analysis is correct. Over the last few years we have come to adopt semantic scaling for this purpose. This technique has served well in psychology (Osgood, Suci and Tannenbaum 1957) and in (socio-)linguistics (e.g. van Bezooijen 1988), and has in fact been applied to the study of intonational meaning at least once before (Uldall 1964). Semantic scaling uses a group of naive subjects as a collective measurement instrument to uncover the semantic dimensionality of a large variety of objects or concepts. Where the intuitions of a single subject, when asked to speculate on the meaning of some word, may be noisy to some degree, the behaviour of a larger group of subjects is likely to be systematic and robust.

**Meaning of *hoor***. In our analysis utterance-final *hoor* is in direct opposition with utterance-final *he*. Both forms claim that there is some personal relationship between speaker and hearer. Both also instruct the hearer to pay particular attention to linguistic material immediately preceding the particle. They are directly opposed to one another, however, in that whereas *he* asks the hearer for some sort of acknowledgement, *hoor* tells the hearer that no acknowledgement is required; he simply has to swallow the message without any reply at all.

In contrast to what was claimed earlier in Kirsner (1991) the speaker-hearer relationship emphasized by *hoor* is not necessarily a friendly one; *hoor* may also express a hostile relationship. The main point, however, is that with utterance-final *hoor*, there is never a neutral or non-existent relationship: *Stikstof is een gas* 'Nitrogen is a gas' could be a factual sentence taken from an encyclopedia; *Stikstof is een gas, hoor* could never be.

**Meaning of boundary tone**. In Dutch, intonation domains (corresponding roughly with clauses or short sentences) usually end either on the low or the high declination line ('t Hart, Collier and Cohen 1990), in autosegmental terminology designated by L (Low tone) and H% (High boundary tone), respectively (e.g. Van den Berg, Gussenhoven and Rietveld 1992). Our analysis of the meaning of the high tone draws on Ohala's (1983) discussion of the ethological basis of certain phonetic features, including high pitch (marking question intonation). Ohala argues that both animals and humans use high pitched sounds, thereby showing or exaggerating their small and harmless physique, in order to express dependence and subservience to the individual they are confronted with. In the animal world this behaviour is claimed avoid bloodshed; in the pragmatics of human communication the high tone is one way of expressing politeness (Brown and Levinson...
We shall assume that the high tone expresses an appeal by the talker on the hearer. The appeal can be interpreted in at least two different ways. Firstly, the speaker may ask for the hearer's continued attention, expressing that there is more to follow. This may be either a linked accent (see below) within the intonational phrase or yet another intonational phrase after the boundary. Secondly, in the case of a question, the speaker may ask the listener for a verbal reply (or a non-verbal compliance with the request). Both interpretations are in line with the abstract meaning of 'appeal to a hierarchically superior hearer', and are compatible with the similar but less general meanings postulated by Keijser (1984). The low boundary tone, on the other hand, does not express appeal or subservience. It rather puts the speaker at a superior or at least equal hierarchical position in the communication process. The speaker indicates that he is finished and/or does not need or want a reaction on the part of the hearer.

Note that there is a partial overlap between the abstract meanings of H% and hoort: both H% and the particle express a non-neutral relationship between speaker and hearer; however, in the case of H% the speaker needs something from the listener, whereas hoor preempts any reaction on the part of the listener. Conversely, there is a complementary overlap between L and hoort in that both boundary and particle indicate that no reaction is required from the hearer. In this case, of course, the difference between L and hoor is that the particle expresses a non-neutral speaker-hearer relationship, which element is absent from L. The consequences of the overlaps in meaning between particle and boundary tone are hard to predict at this juncture; it seems obvious nevertheless that they will interact heavily in the interpretation process.

Meaning of accent linking. A frequent intonation contour that occurs in Dutch is the so-called flat hat ('t Hart et al. 1990). In this contour an accent-lending rise is executed on the pre-final accent in the sentence, followed by an accent-lending fall on the last accent, maintaining high level pitch between the two accents. The contour is often referred to as 'linked accents'. Obviously, when a Speaker initiates this linked contour on the pre-final accent, he must have a clear idea of what the remainder of the sentence is going to be; for instance, he must know that exactly one more accent is to follow before the termination of the sentence. Accent linking presupposes planning. In accordance with this insight, we suggest that the linked contour expresses the absence of surprise; the proposition made in the final part of the sentence is presented as obvious and transparent. The alternative is de-linking: here each accent is implemented as a separate rise-fall configuration (see figure 1). Rise-fall accents can be executed anywhere in the sentence, including pre-final and final position, and require no specific pre-planning. This particular choice of accents is likely to occur when the speaker has not completely decided what he is going to say next; as a result, ending a sentence with two de-linked accents reveals to the listener that the proposition may be less than obvious and contain an element of surprise. These accents are likely to draw the listener's attention more forcefully to the verbal contents of the
utterance, in a way that resembles one of the components of the abstract meaning postulated for *hoor*. As a result, we predict that the presence of *hoor* and de-linking of accents should influence the interpretation process in like fashion, i.e. additive effects rather than interaction.

Research questions. To conclude this introduction let us briefly summarize the questions that underlie the present research:

1. What is the effect of presence versus absence of the sentence-final particle *hoor* on the interpretation of the sentence?
2. What is the effect of a low versus high boundary tone on the interpretation of the sentence?
3. What is the effect of linked versus de-linked accents on the interpretation of the sentence?
4. To what extent do these three factors interact?

2. Method

Materials. A male speaker of Standard Dutch, an experienced phonetician and well-versed in the Dutch intonation grammar, recorded the utterances:

\[
\begin{align*}
\text{Chinees is een toontaal, hoor} & \quad \text{Dertien is een priemgetal, hoor} \\
\text{‘Chinese is a tone language, hear’} & \quad \text{‘Thirteen is a prime number, hear’}
\end{align*}
\]

pronounced with a double rise-fall intonation pattern and with a high boundary tone. These sentences, expressing encyclopedic information only, were chosen since they were thought to be semantically neutral. The utterances were digitally stored (12 bits, 10 kHz, 4.5 kHz LP) and subjected to LPC formant and bandwidth extraction (F1-F5, B1-B5, 25.6 ms window, 10 ms time shift) using the Split-Levinson robust formant determination algorithm (Willems 1987). Fundamental frequency was determined by subharmonic summation (Hermes 1988) followed by automatic pitch tracking. The original pitch contours were stylized to a minimal number of straight lines, as exemplified in figure 1.

Of each utterance a second version was created by removing the final particle *hoor* and copying its high boundary tone onto the preceding word, taking necessary precautions to make the preceding word (*toontaal, priemgetal*) fit the utterance-final position, i.e. lengthening the segments from the stressed vowel onwards by 40% after replacing the last segment by a similar segment excerpted from the same word spoken in prepausal position. Next, four more exemplars were obtained by replacing the stylized double rise-fall contours by flat hat linked accent contours, as indicated in figure 1. Finally, eight more exemplars were made by replacing the utterance-final high boundary tone by a stretch of low declination. In all, 16 utterances were generated in a 2 (basic utterances) * 2 (with/without *hoor*) * 2 (flat hat/double rise-fall contour) * 2 (high/low boundary tone) factorial
design. These utterances were recorded onto two audiotapes in different random orders after LPC-resynthesis and D/A-conversion.

Figure 1: Waveform and three representative stylized pitch contours on test utterance Dertien is een priemgetal, hoor. Top curve: linked accents plus H%; middle curve: linked accents plus L; bottom curve: delinked accents plus L. Note that curves have been displaced by 3 semitones for the sake of clarity.

Subjects and procedure. Two groups of 20 native Dutch students at Leiden University listened to the recordings played to them over loudspeakers in a quiet lecture room. In the first part of the experiment subjects listened to each utterance four times in a row with 2s in between repetitions; after each repetition they rated the utterance along one of four different 7-point speaker scales:

- nagging
- pedantic
- full of oneself
- distant

- satisfied
- docile
- modest
- involved

In the second part of the experiment the 16 utterances were repeated one more time, while subjects indicated for each single utterance how pleasant they thought it would be to be talked to in the fashion exemplified by the utterance:

- unpleasant
- pleasant

3. Results and preliminary conclusions

---

1 The Dutch adjectives actually used in the instructions and answer sheets were, in the same order, zeerderig - tevreden, bekerend - volgzaam, zelfingenomen - bescheiden and afstandelijk - betrokken.

2 The Dutch adjectives used here were onprettig - prettig.
Reliability. The five semantic scales used in our experiment were selected from a much larger set of potentially suitable adjectives in a pilot experiment, such that they optimally differentiated between the various stimulus conditions embodied in our experiment. We therefore have every reason to expect these scales to be reliable in the context of the present experiment. Still, to obviate the possible criticism that the task at hand might be impossible or meaningless for Dutch listeners, we ran a formal check on the statistical reliability of the measurement instrument, using Cronbach’s Alpha as the relevant statistic (Lord and Novick 1968; Nunnally 1978). Alpha can be interpreted as a measure of relative agreement among the listeners, an average correlation coefficient ranging between zero (no agreement) and unity (perfect agreement). The results are as in table 1.

Table 1: Cronbach’s Alpha reliability coefficients for the five semantic scales used in the experiment

<table>
<thead>
<tr>
<th>Scale</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>nagging - satisfied</td>
<td>.95</td>
</tr>
<tr>
<td>pedantic - docile</td>
<td>.87</td>
</tr>
<tr>
<td>full of oneself - modest</td>
<td>.88</td>
</tr>
<tr>
<td>distant - involved</td>
<td>.80</td>
</tr>
<tr>
<td>unpleasant - pleasant</td>
<td>.97</td>
</tr>
</tbody>
</table>

Table 1 shows, quite unequivocally, that our listeners reacted in a highly uniform fashion to the various utterances in the test. We can therefore safely conclude that the subjects gave consistent and systematic judgments to the test utterances in a meaningful task.

Intercorrelation. In the next stage of the data analysis we asked ourselves to what extent each of the five scales contributes independently to the evaluation of the experimental utterances. Table 2 presents the relevant correlation matrix.

Table 2: Correlation matrix for semantic scales used in experiment (*: p<.05).

<table>
<thead>
<tr>
<th>Scale</th>
<th>nagging</th>
<th>pedantic</th>
<th>full of oneself</th>
<th>distant</th>
</tr>
</thead>
<tbody>
<tr>
<td>pedantic</td>
<td>.52*</td>
<td></td>
<td>.92*</td>
<td>-.19</td>
</tr>
<tr>
<td>full of oneself</td>
<td>.62*</td>
<td>.92*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>distant</td>
<td>.34</td>
<td>-.28</td>
<td>-.19</td>
<td></td>
</tr>
<tr>
<td>unpleasant</td>
<td>.96*</td>
<td>.52*</td>
<td>.67*</td>
<td>.34</td>
</tr>
</tbody>
</table>

It appears from table 2 that the scales nagging and unpleasant provide virtually the same information (r=.96). On a common sense basis, it is reasonable that
hearsers would find it unpleasing to be talked to by a person whom they judge to be nagging or whining. Secondly, the pedantic scale runs parallel to the full of oneself scale ($r=.92$), which makes sense insofar as a speaker who is judged as pedantic, lording it over his hearer, will also likely be found to be full of oneself for otherwise he would not lord it over his hearer. Interestingly, the distant scale, finally, seems to be uncorrelated with any of the other scales. In the remainder of this article we shall explicitly treat the five rating scales as organised in three groups, in the way suggested by the correlation matrix.

**Analysis of variance.** After these preliminaries, the data were subjected to separate analyses of variance, one for each scale, with accent linking (linked versus unlinked accents), boundary type (L versus H%), and particle (+/– hoor) as fixed factors, and lexical sentence type (Chinees versus Dertien) as a random factor, with repeated measures over listeners (N=40 per cell). Table 3 presents the results of the analysis broken down by scale and by factor. We also included the only significant interaction found in the analysis: particle by boundary type. Effects or interactions were included in the table only if their significance exceeded the .01 level for at least one of the five scales.

**Accent linking.** The results show, first of all, that - counter to our hypothesis - there is no effect at all due to the way the two accents in our utterances were linked. The accent factor is insignificant on all five of the semantic scales. Either there is no difference in meaning between linked accents (flat hat contour) and unlinked accents (two rise-fall contours) or the differences are such that they cannot be ascertained by the present experimental method and/or choice of scales.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Factor/interaction</th>
<th>accent</th>
<th>boundary</th>
<th>sentence</th>
<th>particle</th>
<th>boun*part</th>
</tr>
</thead>
<tbody>
<tr>
<td>distant</td>
<td></td>
<td>-</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>pedantic</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>satisfied</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>nagging</td>
<td></td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
<tr>
<td>unpleasant</td>
<td></td>
<td>-</td>
<td>-</td>
<td>**</td>
<td>**</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 3:** Summary results of analysis of variance on five semantic scales with accent linking, boundary type, particle and sentence as factors. Only effects and interactions which explain more than 2 percent of the variance ($\eta^2 > 2\%$) have been listed (*: $p<.01$; **: $p<.001$).

**Boundary tone.** Secondly, there is a (relatively small) effect of boundary type for only the distant-involved scale. The effect is plotted in figure 2, where the mean...
judgments are presented for utterances ending in L versus H%:

![Figure 2: Mean 'distant' scale values for utterances ending in L and H%.](image)

![Figure 3: Mean scale values for 'nagging' and 'unpleasant' broken down for Chinees and Dertien utterances.](image)

Stimuli ending in H% are judged to be uttered by a (relatively) more involved talker than utterances ending in L, F(1,39)=30.4 (p<<.001). This finding, of course, supports our postulated meaning for the H% tone. Since H% expresses an appeal made by the talker to the listener (either to pay continued attention or to provide a reply), the speaker is necessarily more involved with the listener than when - ceteris paribus - no appeal is being made.

Sentence type. Thirdly, there is a consistent influence of sentence type, to the effect that, overall, the Chinees utterances are evaluated as (relatively) more nagging and less pleasant than the Dertien utterances, F(1,39)=81.6 (p<<.001) for nagging and F(1,39)=47.8 (p<<.001) for unpleasant. This effect is shown in figure 3, which plots the scale values as a function of sentence type with separate lines for nagging and unpleasant.

It should be reiterated at this point that there is not a single interaction between any of the crucial factors and sentence type, so that this effect has no consequences for the interpretation of the results. If we were to speculate on its cause, however, we would venture that the sentence effect is an artifact of the resynthesis technique used. Informal listening reveals a distinct nasal twang pervading the Chinees utterances which is not noticeable in the Dertien utterances, even though the number of nasal consonants is the same across the two sentence types. Be this as it may, we suggest that the hypernasality of the Chinees utterances contributes to a more strongly felt nagging characteristic, and hence greater unpleasant-ness.

Particle. Now turning to the crucial particle variable, we observe large effects of the presence of hoor on four out of the five semantic scales. Figure 4 plots these scale values as a function of absence versus presence of the final particle hoor,
with each semantic scale identified by a separate line.

Tagging the sentence by *hoor* is evaluated by the listeners as suggesting a (relatively) more *nagging*, more *pedantic* and more *full of oneself* speaker, and is judged to be a (relatively) less *pleasant* style of delivery to listen to. The effects are very strong, \( F(1,39)=141.5 \) \((p<.001)\) for *nagging*, \( F(1,39)=14.4 \) \((p=.001)\) for *pedantic*, \( F(1,39)=15.1 \) \((p<.001)\) for *full of oneself*, and \( F(1,39)=120.9 \) \((p<.001)\) for *unpleasant*. In the case of *nagging* and *unpleasant* the effect of adding *hoor* amounts to a drop along the 7-point semantic scale of 1.4 and 1.9 points, respectively. Such large effects can only be found if the subjects perform their evaluation task very consistently and uniformly.

We assumed that by suffixing *hoor* to the sentence, the speaker draws explicit attention to the existence of a personal relationship between himself and the addressee. In many contexts, such emphasis on the personal bond between speaker and hearer would be evaluated favourably by listeners. However, figure 4 shows unambiguously that with encyclopedia sentences communicating purely factual information, the personal relationship between speaker and hearer is evaluated as more negative in the presence of *hoor*. We shall consider the mechanism behind the negative interpretation of *hoor*-sentences in the following section. For now it suffices to conclude that the departure from neutrality shown in figure 4 is entirely in accord with the abstract meaning assumed for *hoor*.

**Particle by boundary tone interaction.** We now come to the issue that lies at the core of the present research. Since both linguists and naive informants persistently claimed that they were unable to comment on the interpretation of *hoor* in abstraction of a specific intonation pattern, we set up the present experiment as a fully factorial design in which presence versus absence of *hoor* was cross-combined with the high and low boundary tone. For both *hoor* and H% we set up abstract meanings that involve a non-neutral or personal relationship between speaker and hearer. However, *hoor* differs from H% in that the latter makes an appeal to the listener, asking for some sort of reaction (be it continued attendance to the speaker or the issuing of a reply) whereas *hoor* does not ask for (in fact prohibits) any form of reaction or acknowledgement on the part of the listener. It follows from these considerations that there should be a strong interaction between particle and boundary tone. Table 3 reveals that such interac-
tions do indeed exist for two out of the five semantic scales, viz. the pedantic and 
full of oneself scales, which are (cf. table 2) highly intercorrelated. The relevant 
breakdowns are presented in figures 5 and 6, respectively, where the scale values 
have been plotted for sentences with and without hoor and with L versus H%.

Figure 5: Mean scale values for 'pedantic' broken down by presence versus absence of hoor and 
by type of boundary tone.

Figure 6: Mean scale values for 'full of oneself' broken down by presence versus absence of hoor 
and by type of boundary tone.

The configuration of means reveals that tagging the sentence by hoor has no 
effect on the evaluation scales when the utterance ends in a low boundary tone. 
However, when the utterance contains the final high boundary, adding hoor causes 
the speaker to be perceived as (relatively) more full of oneself and pedantic, in 
both cases by 1.4 points. These interactions are very strongly significant, F(1,39)= 
53.0 (p<.001) for full of oneself, and F(1,39)= 39.7 (p<.001) for pedantic.

We conclude that we were correct in predicting that hoor and boundary tone 
should interact. The discussion of the nature of this interaction depends on the 
mechanism underlying the more negative interpretation of encyclopedia sentences 
containing hoor, to which we now turn.

4. Discussion and final conclusions

Although friendly-unfriendly was not among the final set of attributions selected 
for our experiment, so that this aspect could not be measured directly, our results 
show that tagging hoor to a factual, encyclopedia type sentence is not normally 
interpreted as friendly. The speaker of sentences tagged with hoor was evaluated 
as (relatively) more pedantic, more satisfied, more nagging, and the listeners 
stated that being spoken to in this fashion was more unpleasant. This state of 
affairs would never result if hoor was interpreted as a sign of friendliness. As will
be obvious from our introduction, we no longer take the view that friendliness is a necessary ingredient of *hoor*; rather, we suggest that the meaning of *hoor* involves the expression of a non-neutral relationship between speaker and hearer, but not necessarily a friendly one. The reason for this is the attention focusing property of *hoor*, together with the fact that *hoor* tells the hearer that no confirmation is being requested from him. Now *hoor* is an entirely optional utterance-final particle; explicit use of it in a factual sentence like *Dertien is een priemgetal* suggests that this attention-focusing was somehow necessary, that the hearer in consequence was somehow remiss, that he was unaware of or did not know the factual information in question. Furthermore, because the speaker explicitly tells the hearer by choosing *hoor* (rather than *he*) that the speaker does not need and does not want the hearer’s confirmation, the relationship between speaker and hearer can easily be taken by the hearer to be asymmetrical rather than symmetrical, with the speaker knowing more than the hearer, and thereby rubbing in the fact that the hearer had to be reminded of about something. When such sentences are interpreted in a null-context, as was the case in the present study, there is nothing extra to cue the hearer that the reminder in question is intended in a friendly fashion. The resulting overall effect is negative for the hearer insofar as it is usually unpleasant to be characterized as unaware, ignorant, and in debt to a superior speaker.

Why, then, is the (relatively) negative evaluation of factual *hoor* sentences potentiated with H% and inhibited with L? We suggest that this behaviour is the only way the listener can resolve the clashing meanings of *hoor* and the high boundary tone. H% can either be interpreted as a request for continued attention or for a reply. Since *hoor* preempts an overt reaction on the part of the listener, the listener will rule out the reading of H% as a request for a reply. What remains is the alternative reading of H%, viz. a request for continued attention. However, nothing follows after the encyclopedic sentence, so that the hearer will take this as an invitation to fill in the unspoken reproach for himself. When *hoor* is spoken with a low boundary tone, the hearer feels no invitation to finish the unspoken criticism for himself.

To conclude this article, we argue that the results of our scaling experiment support the view that the abstract hint-like meanings postulated for both the particle *hoor* and the high boundary tone are essentially correct. It is also clear now why native speakers of Dutch generally claim that they cannot really interpret an utterance containing *hoor* in the absence of a concrete anchoring intonation pattern (either a real one or an imaginary one, i.e., one which they themselves provide). Our results also strengthen the view that particle meaning and intonational meaning - similar in their relative imprecision when compared to lexical meanings - can to some extent be traded, not only across languages but even within a single language. As a final methodological point, we have shown that semantic scaling is a reliable and insightful technique for objectively evaluating hypotheses concerning sparse, abstract or hint-like interactional meanings of
the sort postulated here

5 References

Abraham, W., ed (1986) 'Partikels', TTT Interdisciplinair Tijdschrift voor Taal- en Tekstwetenschap, 6
Lord, F. M. and M. R. Novick (1968) Statistical theories of mental test scores, Addison-Wesley, Reading MA.