

- BARBOSA, Tito Montenegro. *Uma Fundamentação Ontofenomenológica do Direito*. 1991, 122p.
- BIZ, Osvaldo. *Informática e Soberania*. 1988, 172p. O livro historia os caminhos que levaram o Brasil a adotar a reserva de mercado para a Informática até 1992.
- BRASIL, Luiz Antonio de Assis (org). *Contos de Oficina 3*. 1989, 136p. É o terceiro volume de contos produzido pelos alunos da Oficina de Criação Literária do Curso de Pós-Graduação em Letras da PUCRS.
- BRASIL, Luiz Antonio de Assis (org). *Contos de Oficina 4*. 1990, 112p. Antologia de contos das Oficinas de Criação Literária do Curso de Pós-Graduação em Letras da PUCRS.
- BRASIL, Luiz Antonio de Assis (org). *Contos de Oficina 6*. 1991, 120p. Antologia das Oficinas de Criação Literária do Curso de Pós-Graduação em Letras da PUCRS.
- CERQUEIRA, Siomara Vilanova. *Administrando a Mudança Rumo à Criatividade*. 1989, 60p. Propõe fornecer alternativas para uma mudança no sentido de ajudar o professor a administrar melhor sua criatividade e a de seus alunos.
- CLEMETE, Ir. Elvo. *Leitura & Crítica Literária*. 1990, 185p. Coletânea de ensaios do autor abordando a teoria e a prática da Crítica Literária.
- CLOTET, Joaquim e outros. *A Justiça*. 1988, 104p. A obra tem ensaios dos professores U. Zilles, Reinhold A. Ullmann, Francisco de A. Santos, Sírio Lopes Velasco, Edvino A. Rabuske e Joaquim Clotet.
- GRIGS, Dadeus. *A Descoberta Científica de Deus*. 1989, 296p. Numa lógica cerrada do pensamento, o autor movimenta-se desimpedidamente na Biologia, Física, Geologia, Filosofia e Teologia indagando pela transcendência.
- JOVCHELOVITCH, Mariova. *Encontros Dialógicos: uma vivência em Serviço Social*. 1989, 60p. Constitui um instrumento metodológico valioso para o Serviço Social, fundamentando a relação de ajuda no diálogo e na fenomenologia.
- MOTTIN, Antônio. *De Maróstica a Garibaldi: Memória da Imigração Italiana*. 2ª edição. 1990, 163p.
- ULLMANN, Reinhold Alcysio. *Epicuro: o Filósofo da Alegria*. 1988, 100p. O livro resgata a pessoa de Epicuro e seu pensamento filosófico.
- ZILLES, Urbano. *Grabriel Marcel e o Existencialismo*. 1988, 128p. A obra expõe criticamente o pensamento de Grabriel Marcel no contexto das filosofias contemporâneas da existência.
- ZILLES, Urbano. *O Problema do Conhecimento de Deus*. 1989, 68p. Apresenta a abordagem dos diversos caminhos seguidos na filosofia ocidental para chegar ao conhecimento de Deus.

PEDIDOS DIRETAMENTE À:

LIVRARIA EDITORA ACADÊMICA LTDA.

Av. Ipiranga, 6681 - Prédio 09 - Caixa Postal 1429
90 620 - Porto Alegre/RS

O TESTE - C DE LEITURA: UMA VISÃO PANORÂMICA*

Rüdiger Grotjahn
Ruhr - Universität Bochum

Resumo

O presente artigo dá uma visão geral sobre o campo de pesquisa do Teste-C, uma variante do Procedimento Cloze. Primeiro, pormenorizará uma breve referência sobre o Princípio Cloze e a teoria básica da redundância reduzida. Depois, segue uma descrição e discussão crítica dos clássicos princípios de construção do Teste-C. A seguir, serão esboçados o desenvolvimento e a análise do Teste-C, assim como o seu emprego na pesquisa e na prática. Segue uma discussão da confiabilidade, objetividade, economia e validade do Teste-C, em que reside o ponto principal no âmbito da validade. Em contato com estes dados aprofundaremos relativa e minuciosamente a questão, se o Teste-C é (também) um teste de leitura. Neste contexto também serão referidas algumas pesquisas empíricas, que o autor realizou em resposta à questão.

Abstract

Der vorliegende Beitrag gibt einem Überblick über den Forschungsstand zum C-Test - einer Variante des Cloze-Tests. Als erstes wird kurz auf das Cloze Prinzip und die zugrundeliegende Theorie der reduzierten Redundanz eingegangen. Darauf folgt eine Beschreibung und kritische Diskussion des klassischen Konstruktionsprinzips von C-Tests. Danach wird die Entwicklung und Analyse von C-Tests sowie deren Einsatz in Forschung und Praxis skizziert. Es folgt eine Diskussion der Reliabilität, Objektivität, Ökonomie und Validität von C-Tests, Anschlusswobei der Schwerpunkt im Bereich der Validität liegt. Im Anschluss daran wird relativ ausführlich auf die Frage eingegangen, ob der C-Test (auch) ein Lesetest ist. In diesem Zusammenhang werden auch einige empirische Untersuchungen referiert, die der Verfasser zur Beantwortung der Frage durchgeführt hat.

* Comunicação apresentada no III Congresso Internacional de Psicolinguística Aplicada, realizado na Universidade de Toronto (Canadá) de 16 a 21 de julho de 1991. A tradução, a partir do original alemão, foi realizado por Gisela Greve Hoff do Centro de Pesquisas Linguísticas da PUCRS.

1. Introdução

O Teste-C baseia-se em uma variante do Princípio Cloze e, com isso, em suposições teóricas semelhantes. Contudo, em contraposição ao clássico Procedimento Cloze, não são palavras inteiras que são extintas no Teste-C, mas sim parte das palavras, que então são reconstruídas pelos sujeitos (confira mais adiante a rigorosa descrição dos princípios de construção). Desde sua implantação no ano de 1981, o Teste-C foi observado em crescente proporção, não só pelos pesquisadores, como ainda conseguiu a consideração dos aplicadores do teste. O Teste-C é objeto de um sem número de estudos, uma série de trabalhos acadêmicos e algumas dissertações (Stemmer, 1991) e poderia, com isso, ser visto em conjunto como um dos mais novos testes de linguagem exaustivamente pesquisados. É notável, que o Teste-C em contraposição ao Procedimento Cloze é delimitada uma estreita imagem positiva na literatura.

2. O Clássico Procedimento Cloze como Fundamento do Teste-C

Por muito tempo foi válido nada menos do que o Procedimento Cloze como Teste de linguagem, como o mais apropriado instrumento para medida do estado de língua (confira em seguida também Feldmann, Grotjahn & Stemmer, 1986). Abstrai-se desde o emprego do Procedimento Cloze até o diagnóstico de habilidades específicas, a seguir os Procedimentos Cloze são chamados testes integrativos ou holísticos e medem quase globalmente conforme a competência (leitura) em uma L1, L2 ou LE. O Procedimento Cloze é de mais a mais um teste designado *reduced redundancy* (compare além disso Klein-Barley, 1985d, 1985e; Gaies, 1987). A frequência de soluções corretas vale como medida para a capacidade de direção de uma competência do emprego da lingüística de Spolsky (1973) designado "general language proficiency", respectivamente um mecanismo do emprego da lingüística de Oller (1973) designado "pragmatic expectancy grammar". A gramática de expectativa internalizada forma a base para o processamento construtivo, processo antecipatório da linguagem. O Procedimento Cloze compreende após Oller o cunho dessa expectativa da gramática e com isso também, simultaneamente, a competência integral em uma língua (confira também Oller, 1976, 1979, assim como a crítica de Vollmer & Sang, 1980).

Em tempos mais recentes, na verdade são manifestadas críticas construtivas aos Procedimentos Cloze (confira particularmente Alderson,

1978, 1979; Klein-Braley, 1981). Entre outros menciona-se os seguintes pontos críticos:

1) Para obter um número suficiente de itens, um Procedimento Cloze deve ser relativamente longo.

2) Com base no princípio de anulação os Procedimentos Cloze são muitas vezes constituídos por somente um texto mais longo; este, devido à especificidade do texto, pode levar a um prejuízo de sua imparcialidade.

3) Os fatores 'texto', 'cotas de anulação', e 'ponto de partida' influenciam os coeficientes de confiabilidade e de validade.

4) Junto aos exatos métodos de valorização os Procedimentos Cloze são, freqüentemente, muito difíceis mesmo para os competentes falantes da língua materna. A avaliação após a aceitabilidade da solução é, geralmente, em contrapartida, muito demorada e menos confiável.

5) A dificuldade dos Procedimentos Cloze depende da relação de freqüência das palavras de conteúdo e das palavras estruturais anuladas.

6) Como particularmente resultaram as análises secundárias de Klein-Braley (1981), os Procedimentos Cloze são menos confiáveis do que amplamente aceitos.

Como conclusão das críticas Christine Klein-Braley e Raatz (1981) apresentam o Teste-C (confira Raatz & Klein-Braley, 1982).

O Teste-C baseia-se nos mesmos fundamentos do teste teórico e lingüístico-teórico do Procedimento Cloze; porém, diferencia-se no formato do teste. Ele é válido do mesmo modo que o Procedimento Cloze no que se refere à operacionalidade dos conceitos *reduced redundancy testing* e com isso depende da seguinte aceitação: quanto mais competente são os aprendizes de uma língua estrangeira, tanto melhor uso eles podem fazer de sua gramática de expectativa pragmática e da redundância natural de um texto, e tanto melhor eles podem resolver o Teste-C. Além disso, o Teste-C, bem como o Procedimento Cloze, é geralmente considerado como instrumento de medida integrativo para a apreensão bastante global da competência lingüística na L1, L2 e LE.

3. O Clássico Princípio de Construção

3.1 Descrição e Exemplo

Relativo à crítica ao Princípio Cloze, Klein-Braley e Raatz empregavam as exigências a seguir nos novos processos de teste:

"1. Em textos muito curtos, um teste deveria abranger ao menos 100 itens;

2. Junto à alternativa do ponto de partida e da cota de anulação não deveria haver problemas;

3. As palavras anuladas, com respeito à espécie de palavra, deveriam ser uma amostragem representativa absoluta dos textos globais.

4. Se possível, os sujeitos com conhecimentos especiais não deveriam ser preferidos;

5. Somente a valorização exata deveria ser admitida, pelo que o procedimento viria a ser absolutamente objetivo;

6. Competentes native speakers, por exemplo, deveriam saber realizar o teste com correção quase total;

7. Evidentemente, o novo método deveria ser altamente confiável e válido e, ao mesmo tempo, ainda ser relativamente fácil de desenvolver." (Klein-Braley & Raatz, 1982, p. 24)

Klein-Braley e Raatz acreditam ser suficiente, através destas exigências, a escolha do seguinte formato de teste: a fim de alcançar uma maior proporção no teste imparcial, quer dizer, a fim de não preferir sujeitos com conhecimentos especiais (confira a exigência 4), em contraposição ao Procedimento Cloze os Testes-C serão mais curtos (compostos de cerca de 60 à 80 palavras), textos escolhidos com as mais diferentes temáticas (quanto ao princípio de construção confira os procedimentos em Grotjahn, 1987a; Raatz & Klein-Braley, 1985). Os textos escolhidos devem formar de cada vez uma unidade de significado, possuir conteúdo neutro e ser interessante, não conter vocabulário especial, não exigir conhecimento especial, grupo alvo adequado e tanto quanto possível ser autêntico.

A fim de satisfazer as exigências 1 e 2, será escolhido o seguinte princípio de anulação: começando com a segunda palavra da segunda frase, será anulada em cada texto a cada segunda palavra a segunda metade. Palavra com uma só letra e nomes próprios não serão levados em consideração. Palavras unidas por apóstrofe ou hífen como no francês *l'un* ou *celui-ci* serão contadas como uma palavra. Se uma palavra tem n letras e o n é ímpar, então importa o número de letras anuladas $(n + 1)/2$, quer dizer, nas palavras com número ímpar de letras, o número de letras anuladas é uma vez maior do que o número de letras não pagadas. Em cada texto deve ser realizado o número idêntico de anulações. No final do texto deve constar uma curta e incólume parte do texto como contexto para a solução.

Em muitas pesquisas com textos em alemão e em inglês, Klein-Braley e Raatz observam que também realizam-se as exigências 3 e 6 através do formato Teste-C (confira Klein-Braley & Raatz, 1982; Raatz & Klein-Braley, 1983).

Os textos escolhidos são organizados com dificuldade gradual. Um ponto será dado para cada reconstrução correta da palavra original (ou às vezes também para cada reconstrução aceitável). Relativamente ao princípio descrito, um Teste-C é composto na maioria de 4 à 6 textos com 20 à 25 itens

(lacunas). Um texto de instrução¹ encontra-se na descrição 1 (de Feldmann, Grotjahn & Stemmer, 1986, p. 327).

A seguir você encontrará seis diferentes textos curtos, nos quais a cada segunda palavra, uma parte dela é omitida. Por favor, tente completar a parte que falta.

La vida evoluciona. La sociedad, también. Hoy día existen bienes de consumo que la comunidad entiende como imprescindibles para la vida cotidiana doméstica; un frigo ____, el ag ____, corriente, l ____, energía eléc ____, no só ____, son d ____, gran útil ____, sino q ____, su inexistencia ____, impediría e ____, una soci ____, moderna. D ____, mismo mo ____, existen cie ____, productos alimen ____, de l ____, que sie ____, hay u ____, reserva e ____, el ho ____, en to ____, los hog ____. Dentro de este "stock" doméstico a que nos referimos, ocupan lugar preponderante los vinos, aguardientes y licores en proporción variable según el gusto del interesado.

Descrição 1

Teste-C em espanhol: instrução e exemplo de texto

3.2 Alguns Problemas com o Clássico Princípio de Construção

O clássico princípio de construção está permeado de uma série de problemas (dependentes da língua - confira a discussão em Grotjahn, 1987a; Grotjahn, Hohenbleicher & Tönshoff, no prelo; Kokkota, 1988). Quanto ao francês a regra de anulação não apresenta problemas, já que as formas com apóstrofe na palavra serão tratadas como a única palavra, mas na língua italiana, por exemplo, conduzem a problemas com incontáveis erros. Assim devem ser anuladas, por exemplo, em *dell'anno* ou *quell'epoca* correspondentemente à regra clássica *anno* ou *epoca*. Deve ser evidente que estas anulações só possam ser reconstruídas com dificuldade. Uma possível solução do problema consiste em somente às vezes anular a segunda metade de *anno* ou *epoca*. Contudo, isto leva a itens muito fáceis (de mais a mais propostas de solução serão discutidas em Grotjahn, Hohenbleicher & Tönshoff, no prelo).

Um vasto âmbito são os pronomes pessoais enclíticos, para os quais a clássica regra de anulação conduz a problemas, por exemplo no italiano, ou também no português. Assim, por exemplo, são anulados de todos os pronomes pessoais enclíticos em italiano *regalarglielo*, ou em português *del-lho*, o que em compensação pode conduzir a consideráveis problemas de reconstrução, particularmente, quando os referentes dos pronomes igualmente pertencem às palavras anuladas no texto. Por outro lado, uma

1 Frequentemente, é citada na instrução um pequeno exemplo, no qual as soluções corretas estão escritas manualmente. Isto é necessário antes de tudo, quando os sujeitos do formato Teste-C não são dignos de confiança.

possível saída para esta problemática é, simplesmente, anular a segunda metade do pronome enclítico.

Também nas palavras compostas pode haver problemas de construção. Assim, por exemplo, é com *altopiano* no italiano, que é constituído por um adjetivo (*alto*) e por um substantivo (*piano*), o substantivo é anulado. Aconselha-se no caso das construções de dois componentes, em princípio, esquecer pelo menos a primeira letra do segundo componente.

Também nas línguas como o hebraico moderno, o turco ou o japonês, que se distinguem tipológica e nitidamente, até agora, pelo menos nas línguas analisadas na pesquisa relativa ao Teste-C, a saber, o inglês, o alemão e o francês conduzem a problemas do ortodoxo princípio de anulação. Em consequência, Cohen, Segal & Weiss Bar-Siman-Tov (1984, 1985) não anulam na língua hebraica a segunda, mas a primeira metade de cada segunda palavra. À propósito, o mesmo princípio Cleary (1988) aplicou no inglês. Devido a caracteres aglutinadores do turco Baur e Meder (1989, no prelo) partiram não de palavras, mas sim de sílabas e têm anulado continuamente cada terceira sílaba de um texto. Finalmente, Roos (no prelo) discute os distintos princípios de anulação para o japonês, que não partem de palavras, mas sim de caracteres.

Mais adiante, no caso de aprendizes avançados, muitas vezes os Testes-C são muito fáceis e não discriminam o suficiente num âmbito mais alto a proficiência lingüística. Isto é válido pelo menos para o inglês, o francês e o alemão. Uma possibilidade subsiste nisto, tornar o Teste-C mais difícil, somente danificando palavras de conteúdo. Atualmente, Klein-Braley experimenta testes correspondentes entre aprendizes de inglês em Buisburg (Klein-Braley, comunicação pessoal). Contudo, aqui resulta com limitações o problema, que os Testes-C construídos deste maneira - semelhante ao Procedimento Cloze, baseado no chamado "rational deletion procedure" - possivelmente medem parcialmente outra coisa do que o Teste-C construído de acordo com o princípio clássico.

Todavia, visto em conjunto, eu defendo um princípio de construção flexível, que antes de tudo também considera as diferenças tipológicas entre as línguas (confira também a proposta de Kokkoto, 1988).

4. Evolução e Análise do Teste-C

Na evolução do Teste-C pretende-se o melhor possível uma seleção de cerca de 10 textos. Estes devem ser então submetidos a um grupo de falantes de língua materna. Através desta maneira, entre outras, pode ser comprovado se os falantes de língua materna resolvem corretamente os textos seguintes.

O texto correspondente é eliminado quando a frequência de solução monta a menos de 90%. Em línguas como o inglês, o alemão ou o francês parece ser somente por vezes que este caso é extremamente raro. Todavia, frequentemente, o limite de 90% é alcançado, por exemplo, no espanhol (U. Feldmann, informação pessoal).

Pesquisas-piloto com falantes de língua materna têm larga vantagem na identificação de textos muito difíceis, para serem verificadas variantes de solução aceitáveis. Em contraposição ao Procedimento Cloze, as variantes aceitáveis nos Testes-C são em geral muito raras. Particularmente é válido para o inglês e parece - no entanto, com menor clareza - ser válido também para o francês, o italiano ou o espanhol. Se em um texto aparecem muitas variantes, então pode-se também evitá-las e com isso facilitar a avaliação, anular menos letras nas palavras correspondentes.

Após as pesquisas-piloto com os falantes de língua materna, deve-se realizar também pré-testes com grupos de aprendizes, para os quais o teste é destinado. Estas pesquisas prévias servem para a avaliação da dificuldade de textos isolados, bem como a dificuldade e a confiabilidade dos testes globais. De mais a mais, é necessário considerar que os itens (lacunas) isolados dentro de um texto dependem uns dos outros, e que, por isso, é esgotada a avaliação da confiabilidade de cada texto como superitem. Com base no valor total para cada texto, são então calculados os coeficientes Alfa de Cronbach ou Lambda de Guttman (confira Grotjahn, 1987a). Em menos de 10 textos isto é rapidamente realizado sem computador. Nesta ocasião também é verificada a seletividade de cada superitem (textos). Textos com uma diminuta seletividade e com colaboração negativa, ou somente pequena para a confiabilidade, são então eliminados.

Contudo, para a pesquisa da dimensionabilidade dos Testes-C, é apenas condicionalmente útil a análise de seletividade. Uma base adequada fornece o chamado modelo clássico de teste latente-aditivo de Moosbrugger & Müller (1981, 1982). Como este modelo pode ser aplicado ao Teste-C, por exemplo, é mostrado em Grotjahn (1987a) ou Raatz (1985 a 1985b).

Finalmente, ainda é necessário chamar a atenção de que (extensas) pesquisas-piloto são, na verdade, muito necessárias, que, todavia, em muitas pesquisas também pela primeira vez Testes-C estabelecidos contaram com resultados admiráveis. Isto pode ser interpretado como uma referência à "Robustês" do Princípio do Teste-C.

Até agora nas pesquisas do Teste-C eram medidos predominantemente os níveis da linguagem com os seguintes três grupos: a língua materna de crianças no inglês e no alemão; o conhecimento da língua estrangeira dos aprendizes, universitários ou estudantes de cursos de línguas; bem como o conhecimento de uma segunda língua, por exemplo, dos trabalhadores estrangeiros na República Federal da Alemanha. Até aqui, entre outras, eram consideradas as seguintes línguas (seguindo a ordem alfabética): Alemão, Chinês, Croata (Sérvio), Espanhol, Finlandês, Francês, Grego, Húngaro, Inglês, Italiano, Japonês, Neolandês, Novo Hebráico, Russo e Turco. Em todas as línguas mencionadas, o Princípio do Teste-C mostra-se aplicável, em princípio.

Fora dos contextos de pesquisa, o Teste-C foi instituído, até aqui, entre outras, nas medidas seguintes (eu me refiro à República Federal da Alemanha):

- como teste de ingresso universitário para a língua inglesa e francesa, por exemplo, em Duisburg e em Bochum;

- no âmbito do "Concurso Federal de Línguas Estrangeiras"²ⁿ para mais de 10 línguas (veja, por exemplo, Raatz & Klein-Braley, 1989; Grotjahn, 1989);

- de Bolten (no prelo) como teste de ingresso em cursos de alemão como língua estrangeira;

- de Jakschik (no prelo) no âmbito do Serviço Psicológico das delegacias regionais de trabalho como ajuda decisiva com respeito ao seguimento dos estudos em alemão;

- de Grönwoldt (1985) como teste de ingresso para cursos de alemão com estrangeiros adultos no Eurocentro em Colônia.

Mais adiante, para o 3º e 4º ano primário, há em preparação uma padronizada versão apropriada de um Teste-C para o alemão como língua-mãe (Autores: Ulrich Raatz e Christine Klein-Braley).

Na maioria das pesquisas, o Teste-C mostra-se altamente confiável (em parte, monta a mais de 0,9 o coeficiente Alfa de Cronbach). Isto é surpreendente, quando muitas das versões pesquisadas do Teste-C foram instituídas pela primeira vez.

Ao lado de uma comparação altamente confiável, o Teste-C possui também uma alta objetividade de realização e permite - então, pelo menos, quando soluções isoladas exatas ou soluções aceitáveis anteriormente pré-fixadas foram avaliadas como "corretas" - uma avaliação absolutamente objetiva.

Mais adiante, as pesquisas até aqui realizadas mostram que o Teste-C é um instrumento relativamente econômico, quando comparado com muitos outros. O esforço de desenvolvimento de um Teste-C é ínfimo, por exemplo, comparado com o de um teste de múltipla escolha. A realização de um Teste-C com cinco textos dura menos de 30 min., e o tempo de avaliação monta de 1 à 2 minutos por texto e por sujeito (conforme o método de avaliação).

7. Validade

7.1 Pesquisa de Correlação Estatística

Conforme já foi mencionado, a interpretação em geral será defendida, para que o Teste-C igualmente ao Princípio Cloze represente um integral instrumento de medida para a compreensão global da competência lingüística na L1, L2 e LE. Esta interpretação apoia-se, entre outras, em parte sobre surpreendentemente altas correlações do Teste-C com diferentes critérios externos, como são, por exemplo, as notas escolares, o parecer do professor sobre o padrão lingüístico dos alunos, bem como, com testes lingüísticos padronizados (compare as visões gerais em Klein-Braley & Raatz, 1984; Raatz, 1985c) sobre pesquisa de validade fatoriais (confira, por exemplo, Grotjahn, 1987a, 1987b; Grotjahn & Stemmer, 1985; Raatz & Klein-Braley, 1983; Raatz, 1984, 1985d), bem como sobre hipóteses empiricamente apoiadas para a validade de construto (confira, por exemplo, Klein-Braley, 1985a, 1985b).

Correspondentemente, por exemplo, Klein-Braley (1985b) chega à seguinte conclusão após uma revisão de validade de construto de Teste-C na

2 obs.: a nota número 2 não apareceu no original.

língua-mãe e língua estrangeira, com auxílio de uma rede nomológica na forma de previsões baseadas na teoria sobre correlações entre a operacionalização de construtos envolvidos (confira também Klein-Braley, 1985a, p. 101):

"I consider that the results presented here are sufficient evidence to substantiate the claim that C-Tests are construct valid tests, and that the construct they are operationalizations of could be called 'general language proficiency.'" (p. 65)

Mais adiante será diferenciada esta interpretação, entre outras, por Bauer & Meder (1989) em relação a suas pesquisas com a interdependência da habilidade na L1 e na L2 dos filhos de imigrantes na República Federal da Alemanha (confira também Bauer & Meder, no prelo). Os dois autores, com limitações, constataam:

"O Teste-C examina abertamente capacidades lingüísticas, que só são adquiridas e treinadas na escola e baseada na leitura e na escrita. Quando se distingue entre uma capacidade de comunicação lingüística-oral geral e uma capacidade desenvolvida no contexto escolar, com capacidade lingüística combinada, diferenciando entre si leitura e escrita - como sucede no âmbito da hipótese de interdependência - então é evidente, que o Teste-C em um sentido global somente examina estas capacidades lingüísticas secundárias." (Bauer & Meder, 1989, p. 123)

De acordo com Bauer e Meder, o Teste-C verifica, por conseguinte, a capacidade lingüística secundária combinada com a leitura e a escrita. O domínio da linguagem comunicativo-interativo não pode ser alcançado de acordo com a interpretação dos dois autores.

A capacidade lingüística secundária, no sentido de Bauer e Meder, será delineada na literatura anglo-americana sob o aval de Cummins (p. ex., 1983) também como "Cognitive Academic Language Proficiency (CALP)" (confira Bauer & Meder, 1989, p. 120) e será diferenciada em "Basic Personal Communicative Skills (BICS)". Isto significa que o Teste-C (primariamente) deveria ser visto como um Teste-CALP (confira também Grotjahn & Tönshoff, no prelo).

7.2 Validade Introspectiva

Uma imensa parte das pesquisas, com respeito à validade dos Testes-C, limita-se à análise de correlações dos valores do Teste-C com diversos critérios externos. Embora pesquisas de correlação estatística possam prestar uma importante contribuição para a averiguação da validade do Teste-C, com restrições não podem responder satisfatoriamente à questão de validade somente com o auxílio de pesquisas de correlação estatística (para uma explicação mais enriquecedora confira Grotjahn, 1986 e Bachman, 1990, cap. 7).

É necessário, para a valorização dos testes, também uma análise de processos mentais, que decorrem da realização dos testes com os sujeitos. Uma exigência análoga pela mudança da habitual prática de pesquisa orientada de correlação estatística já foi levantada por Vollmer (1981; 1983), entre outros, que constatou:

"What we suggest in the application of alternative research strategies... We will never really understand what correlations between tests of different skills mean, what they signify, and why some are higher than others - unless we better understand and are able to model more precisely the cognitive potentials and take specific operations on which performance in the various language tests depends." (1981, p. 115)

Entretanto, há uma série de pesquisas para os processos mentais, que ocorrem no aprendiz durante a resolução de um Teste-C. Conclui-se do trabalho de Cohen, Segal & Weiss Barsiman-Tov (1984; 1985), que se trata, pelo que eu sabia, exclusivamente de pesquisas, que foram realizadas no âmbito de um projeto de pesquisa na Universidade de Bochum no Ruhr³.

Neste projeto tentamos pela primeira vez em 1983 entender, com o auxílio dos métodos do Pensar Alto e de entrevistas retrospectivas, os processos mentais em desenvolvimento durante a realização de um Teste-C (para o emprego do método introspectivo na pesquisa sobre os testes confira também Cohen, 1984a). Esta experiência de construção, por nós escolhida, pode ser caracterizada da seguinte maneira (confira a descrição detalhada em Grotjahn, 1987c):

Aos sujeitos - trata-se predominantemente de estudantes do sexo feminino e masculino da Universidade de Bochum no Ruhr - eram apresentados de cada vez os textos do Teste-C a serem realizados nos experimentos isolados, três textos em francês, quatro em espanhol, dispostos na ordem crescente de dificuldade. A fim de assegurar que os sujeitos entenderam completamente a tarefa por resolver, eles serão informados relativa e minuciosamente sobre o objetivo da pesquisa. Logo a seguir, uma descrição por escrito é apresentada ao sujeito. Nisto o sujeito é instruído, por exemplo, a dizer tudo o que ele/ela pensava ou sentia, quando preenchia as lacunas do Teste-C. Mais adiante, é executada uma curta gravação como ilustração do Pensar Alto durante a realização de um Teste-C. Finalmente, é apresentado ao sujeito um curto texto do Teste-C como exercício. Após esta fase de instrução, o sujeito é deixado sozinho para realizar o texto do Teste-C. O Pensar Alto é gravado.

3 Direção do projeto: Friedrich Dening, Rüdiger Grotjahn; Colaboradores: Ute Feldmann, Brigitte Stemmer. O projeto foi subvencionado, por vezes, pela Deutschen Forschungsgemeinschaft. Os resultados do projeto estão documentados, entre outros, em Feldmann, Grotjahn e Stemmer (1986), Feldmann & Stemmer (1987), Grotjahn & Stemmer (1985), Stemmer (1991), bem como em Stemmer (no prelo)

Depois o sujeito e o pesquisador ouvem juntos a gravação da fase do Pensar Alto. Nesta fase retrospectiva os sujeitos poderiam espontaneamente tomar posição com respeito às suas exteriorizações e, responder às perguntas do pesquisador, relativas às suas manifestações. Deste modo, foram levantados dados de 30 sujeitos para o francês e de 70 sujeitos para o espanhol. Os dados para o espanhol, em parte, apresentam-se transcritos por escrito. Os dados do francês estão, em sua totalidade, transcritos e avaliados.

Para a análise dos dados, Stemmer (1991) desenvolveu um modelo de análise específico, que se compõe de três partes: 1) uma análise do trabalho teórica sob inclusão de uma análise de texto proposicional; 2) uma análise empírica da solução de Teste-C; e) uma análise dos relatos orais. Sob estes fundamentos foi verificado, entre outros, uma lista de frequência com diferentes estratégias de solução. Stemmer (no prelo) menciona como importante resultado de suas pesquisas, que estão amplamente documentados em Stemmer (1991), a limitação do sujeito na descoberta de solução ao ambiente próximo ao item: "The strategic behavior of the subjects was predominantly directed towards the immediate co-text of the item." Com respeito a este resultado ainda voltarei a falar posteriormente.

8. O Teste-C é um teste de leitura?

8.1 Alguns Resultados da Literatura

Semelhante a Klein-Braley (1985a; 1985b) Cohen, Segal e Weiss Bar-Siman-Tov (1985, p. 127) chegam à conclusão, na sua revisão do Teste-C em hebraico, de que o Teste-C mede a competência lingüística geral: "It - the Hebrew C-test - appears to be both a reliable and valid measure of general language proficiency". Contudo, eles também mostram um outro ponto do Teste-C como "test of reading ability" (p. 125). Mesmo quando se trata de uma rotulação casual, do que de uma afirmação decidida da validade de construto, então, é levantada uma importante questão (confira a seguir Grotjahn & Tönshoff, no prelo): O Teste-C é (também) um teste de leitura e ele poderá ser inserido como tal para testar a compreensão da leitura?

Klein-Braley (1982, p. 59) parece querer responder com um "não" mais do que depressa a esta questão, quando ela afirma com respeito ao Teste-C: "...we could not claim that these tests can be used for measuring reading comprehension." Em comparação, Faust (1986) acredita poder responder a questão apresentada com um inequívoco "sim". Ele escreve:

"Uma possibilidade de aplicação [sic], já inteiramente aplicável do Teste-C, reside na testagem rápida da compreensão de leitura. Tradicionalmente, deve-se fazer perguntas sobre um texto e este deve ser respondido por estudantes isoladamente. Através deste procedimento oral escapa muita informação, já que nem todos podem responder às perguntas... Porém, com o uso de um Teste-C torna-se objetivamente claro, se existem defeitos na compreensão do texto ou não. Infelizmente é muito difícil que a extensão do defeito seja claramente quantificada."

Na verdade, eu creio que até aqui o Teste-C pode ser indicado como 'teste de leitura', como há sem dúvida uma série de processos mentais, que de maneira uniforme ou semelhante não só acabam na realização do Teste-C (confira, além disso, Grotjahn, 1987b). Este deveria ser espontaneamente compreensível também para o nível de plausibilidade: O Teste-C é definitivamente um teste escrito, para cuja realização devem ser 'lidas' as palavras mutiladas e as intactas. Algo que ainda não foi o suficientemente esclarecido é, em que dimensão ocorrem os pontos comuns, e quão significativos são os processos específicos de leitura e os processos específicos do Teste-C.

Porém, em minha opinião, a recomendação de instituir o Teste-C para perceber rapidamente a compreensão da leitura, com certeza, é precipitada e provavelmente injustificada. Até aqui faltam pesquisas para a problemática. Para outros, os resultados apontam a seguir também para os estudos-piloto descritos, que os Testes-C quase nem são indicadores válidos para dificuldades na compreensão de texto.

Da literatura para o Teste-C, com respeito às questões, são formulados os seguintes resultados empíricos de interesse:

- Raatz (1984, p. 138) encontrou na pesquisa com alunos alemães uma correlação de .43 entre o Teste-C da língua materna e um de questões de múltipla escolha para dois textos de leitura na língua materna composto de teste de compreensão de leitura padronizado (subteste do "DTD - Diagnostischer Test Deutsch". - confira também Raatz, 1985d, p. 49).
- Cohen, Segal e Weiss Bar-Siman-Tov (1985, p. 124) verificaram uma correlação de .69 com aprendizes adultos de hebraico, entre um Teste-C hebraico e um de questões de múltipla escolha composto de testes de compreensão de leitura para jovens falantes da língua hebraica.
- Grotjahn (1987b, p. 236) encontrou uma correlação de .29 junto a alemães calouros do francês e um de 13 questões de múltipla escolha, de dois textos de francês composto de testes de compreensão de leitura.

Na avaliação destes e de outros resultados, deve-se levar em consideração, que os testes de leitura utilizados para a validação em parte não são muito confiáveis, e que, por isso, em alguns casos talvez é superavaliada a 'verdade' da validação de critérios⁴.

Mas mesmo quando se fala de correlações de critérios um pouco mais altas, as correlações encontradas na literatura, entre testes de leitura padronizados e o Teste-C, quase não falam em favor da visão de Faust (1986), que o Teste-C examina em primeiro lugar a compreensão de leitura.

Mais adiante, pronunciam-se contra a suposição de Faust também os resultados de muitas pesquisas, entre os quais foram colocados o Teste-C e baterias de testes para a avaliação de ingresso em cursos de língua estrangeira (inglês, alemão, francês, hebraico - confira, p. ex.: Chapelle & Abraham, 1990; Cohen, Segal & Weiss Bar-Siman-Tov, 1985; Grönwoldt, 1985; Grotjahn, 1987a, 1987b; Klein-Braley, 1985c; Raatz, 1982). As correlações entre os resultados gerais dos testes (soma dos resultados do subteste) as baterias de teste em regra foram muito altas (em parte, maiores do que 0,8), que se pode partir do princípio de que as baterias de teste discrete-point, com seus diferentes subtestes, basicamente medem o mesmo como os respectivos Testes-C.

Estes fatos também serão apoiados através de pesquisas fatoriais e analíticas e, através de pesquisas com o clássico modelo de teste latente-aditivo, em cujos resultados do Teste-C produziram uma estrutura unidimensional nos resultados das baterias de teste de línguas estrangeiras (confira p. ex.: Raatz, 1982, 1984, 1985c, 1985d; Grotjahn, 1987a, 1987b).

Com isso as correlações, sobretudo com os subtestes gramaticais das baterias de teste, foram, em parte, inequivocamente mais altas do que a correlação com respectivos testes de leitura⁵.

Finalmente, pronunciam-se contra a suposição de Faust também as correlações altas a médias obtidas para diferentes línguas e populações, entre os valores do Teste-C e as notas escolares em disciplinas de língua, ou seja, avaliações (rankings) globais da competência linguística dos participantes

do Teste-C através de seus respectivos professores (confira, p. ex., Raatz, 1984, 1985c, 1985d). Quase não se pode partir do princípio de que as notas correspondentes e os conceitos dependem, em primeiro lugar, de uma avaliação da capacidade de leitura dos alunos.

Porém, vista em conjunto, a interpretação dos mencionados resultados estatisticamente correlacionados é problemática. Assim, por exemplo, deve ser questionado, se e em que medidas os testes utilizados para a validação realmente medem a compreensão da leitura. Além disso, como já foi mencionado, a validação estatisticamente correlacionada apresenta problemas fundamentais.

8.2 Procedimento para a Medição da Compreensão de Leitura

O construto de 'reading ability' (competência de leitura) pode ser determinado de duas maneiras: 1) sobre o nível de processo, quer dizer, sobre o processo em desenvolvimento no leitor, processo de (re)construção do sentido do texto; 2) sobre o nível de produto, quer dizer, sobre a compreensão de texto obtida posteriormente à leitura (confira Grotjahn, 1987b). Procedimentos de medida como exercícios de múltipla escolha, questões em aberto para importantes conteúdos do texto, 'summary' (resumo do texto) ou 'free recall', quer dizer reprodução oral e escrita do conteúdo do texto após o processo de leitura bem sucedido, sem 'probing' específico (confira Kluwe, 1988, p. 380), são, em sua opinião, os testes de compreensão de leitura, quer dizer, testes que devem medir o resultado do processo de compreensão de leitura (confira Johnston, 1984; em 'recall', ou seja, 'summary', confira especialmente Bernhardt, 1986; Lee, 1986).

Com isto resulta o seguinte problema: os processos básicos podem ser encerrados no nível de produto de maneira válida? Ou dito de outra forma: A compreensão de texto é um indicador adequado para o processo de compreensão de leitura? O resultado de dois produtos semelhantes pode ter estratégias e operações muito diferentes. Com respeito a padronizados testes de leitura de múltipla escolha, Langer (1987) chega à seguinte conclusão:

"Standardized reading tests do not measure the processes involved in the construction of meaning from a text nor do they evaluate an individual's ability to manage these processes." (p. 243)

Os processos introspectivos de levantamento de dados, que têm obtido grande importância na pesquisa do processo de leitura na língua estrangeira, partem precisamente deste ponto: eles devem transmitir conhecimentos decisivos na dimensão processual de retenção da leitura⁶.

4 Assim atinge, por exemplo, .57 a confiabilidade nos testes de leitura aplicados por Grotjahn (1987b). Avalia-se para uma simples correlação obtida de .29, somente a imperfeita confiabilidade dos critérios de correção de considerada redução (confira Lienert, 1969, p. 298), elevando-se o coeficiente de validação de .29 para .38.

5 Assim são, p. ex., as correlações entre os Testes-C e os subtestes gramaticais, em parte maiores do que .70 em Grotjahn (1987b). Em Cohen, Segal & Weiss Bar-Siman-Tov (1985) monta até .87 a correlação entre os testes gramaticais e o Teste-C. Chapelle e Abraham (1990) relatam as seguintes correlações corrigidas de redução entre um Teste-C em inglês e cinco testes maiores: .84 (vocabulário); .64 (escrita); .60 (leitura); .47 (compreensão); .40 ("Group Embedded Figures Test" (GEFT)). O GEFT é um teste de independência de campo. Os quatro testes restantes são testes de múltipla escolha e parte do "Iowa State University English Placement Test". Conforme Chapelle e Abraham (1990, p. 130) o teste escrito é em primeiro lugar um teste gramatical na interpretação de Bachman (1990).

Mais adiante, perguntamos se os procedimentos de medida mencionados medem com bastante validade o produto final do processo de compreensão de texto. Na tentativa de responder a pergunta, coloca-se o seguinte problema ontológico característico da pesquisa de leitura: Diferentemente da escrita, o produto da leitura é puramente mental, ou seja, não é acessível a uma observação direta. O produto existe apenas na cabeça dos leitores. Primeiro através do processo adicional da lembrança, quer dizer, a reconstrução (ativa) do conteúdo de consciência conclui-se diretamente na compreensão do texto.

Procedimentos como questões seletivas, questões em aberto ou 'recall' serão instituídos na língua estrangeira para a leitura de apreensão de sentido; por conseguinte, são esperadas distorções numerosas. Aplica-se, entre outras, as seguintes limitações fundamentais:

- Os procedimentos mencionados medem 'off-line', separado do verdadeiro processo de leitura. Os procedimentos apresentam exercícios propriamente criativos para o organizador, visto que, depois da leitura resultam processos adicionalmente mentais (veja acima). Na verdade, os Testes-C teriam a vantagem de medir 'on-line'. Porém, o processo de recepção propriamente parece ser modificado através de processos específicos do Teste-C, que transcorrem simultaneamente.
- Os resultados dos testes dependem, grandemente, do resultado de lembranças e do modo de armazenamento. Há numerosa documentação para isto na literatura, que, com inferências elaborativas, devem ser calculadas durante o 'recall', quer dizer, com divergências perante o texto original, por exemplo, através da (re)construção de pontos de vista de plausibilidade. Esquemas sobrepostos ativados são 'recheados' de microposições próximas, sem que estes mesmos sejam lembrados isoladamente (para informação de inferência na reprodução de texto, confira, por exemplo, Graesser & Bower, 1990; Rickheit & Strohner, 1985, 1989, 1990). Sobretudo, no caso de ausência de concordâncias entre informação existente, podem surgir traços fortemente idiossincráticos na 'lembrança' da (re)construção do conteúdo do texto, que aumenta a concordância (confira também Ballstaedt, Mandl, Schnotz & Tergan, 1981, p. 95; Hermann & Hoppe-Graff, 1988, p. 289).

6 Para o uso de procedimento introspectivo na pesquisa de leitura (língua estrangeira) confira, p. ex., Afflerbach & Johnston (1984); Block (1986); Cohen (1984b, 1986, 1987); Ericsson (1988); Faech & Kasper (1987); Garner (1987); Hosenfeld (1984); Olson, Duffy & Mack (1984); Sternglass (1988).

Mais adiante, desempenham um papel no 'recall', também, as diferenças individuais, por exemplo, na inteligência ou na escola de introversão (confira, p. ex., Pléh, Oláh & Zétényi, 1988) >

Finalmente, também deve-se considerar simples problemas de lembrança (recall). Assim, a lembrança pode não ser possível, não obstante, a compreensão de texto bem sucedida, por exemplo, por razões emocional-afetiva momentânea.

Estes fatos são considerados na avaliação da pesquisa descrita de Grotjahn & Tönshoff (no prelo) na seção 8.3.

Mediante questões com respeito ao texto, será provido ao leitor com input adicional, que influencia a compreensão de texto, ou seja, a lembrança. Porém, esta interação adicional será evitada no 'recall'. Além disso, nas questões de teste, freqüentemente, não se separa suficientemente a capacidade de leitura e o conhecimento de base específico da leitura (confira, entre outros, Steffenson & Joag-Dev, 1984, p. 51). Se são apresentadas questões formuladas na língua estrangeira, muitas vezes, não se pode resolvê-las inequivocamente, já que problemas de respostas são atribuídos à questão de compreensão falha do texto ou à compreensão insuficiente da pergunta.

Se são exigidas respostas na língua estrangeira, sobre o texto, ou também um 'recall' na língua estrangeira, pode-se esperar problemas adicionais na habilidade basicamente falha e produtiva na língua estrangeira (confira Wolff, 1987, p. 325). Contudo, o problema da habilidade eventualmente insuficiente e produtiva apresenta-se também no 'recall' na língua materna (confira além disso, entre outros, o resultado empírico de Johns, 1985).

Também este fato será ponderado na pesquisa descrita na apreciação da seção 8.3.

Sintomática, para a problemática de perguntas de compreensão de múltipla escolha, é a seguinte experiência de um leitor, que Candlin aponta no prefácio de Alderson e Urquhart (1984): "I got 100% on the questions but I didn't understand the passage" (p. XI). Este resultado anedótico será apoiado através de um grande número de pesquisas empíricas para o chamado "passage dependency" das questões de compreensão de texto (confira, p. ex., Johnston, 1984, p. 154ff.; Tuinman, 1973/1974; Weaver, Bickeley & Ford, 1969).

8.3 As Pesquisas de Grotjahn e Tönshoff (no prelo)

Para a revisão das questões, se o Teste-C (também) mede compreensão de leitura, foram realizados por Grotjahn e Tönshoff (no prelo), assim como por Grotjahn (no prelo), uma série de estudos-piloto. Por razões de espaço, eu posso descrever meramente a ordem das pesquisas e os resultados

principais no âmbito deste artigo. O leitor interessado pode referir-se aos trabalhos mencionados.

8.3.1 Projeto

Os estudos-piloto realizados por Grotjahn e Tönshoff (no prelo) tinham o projeto geral seguinte: Cada vez seria aplicado um texto do Teste-C de francês e de italiano respectivamente aos aprendizes de francês e de italiano com diferentes níveis de proficiência. Todas as pesquisas seriam realizadas com estudantes da Universidade de Bochum ou Ruhr. Depois, seriam preenchidas as lacunas, sem que isso tivesse sido notificado, recolhidos os textos e solicitado aos alunos a escreverem o conteúdo do texto trabalhado em alemão, quer dizer, foi empregado o método *immediate free recall*. Subseqüentemente, os textos não mutilados seriam apresentados aos estudantes - com o desafio de traduzi-los literalmente o quanto possível.

Desta maneira, nós conseguimos três conjuntos de dados:

- 1) a solução do Teste-C;
- 2) o relatório *immediate free recall*;
- 3) a tradução literalmente dos textos (integrais).

Com respeito ao projeto geral, é apontado o seguinte:

- 1) Somente um único texto do Teste-C foi apresentado aos alunos, o qual, contudo, era nitidamente mais longo do que os 50 à 61 itens usuais. Com vistas ao 'recall' seguinte foi eliminada a combinação normalmente prevista para o Teste-C de vários textos curtos de diversos conteúdos.
- 2) No projeto em questão, o 'recall' empregado não é instrumento de diagnóstico independente da aplicação do Teste-C, a respeito da competência local da língua estrangeira. O 'recall' aqui é muito mais dirigido à aplicação do Teste-C. Isoladamente, trata-se da questão segundo o modo da relação entre os resultados de Teste-C e a proporção da exploração do sentido do texto no decorrer do trabalho. Ou dito de outro modo: o 'recall' deve indicar em que extensão os processos 'mais elevados' de elaboração e constituição de sentido, centrais para o processo de leitura de compreensão de sentido, também são eficazes na aplicação do Teste-C.

Foram usados três textos:

- 1) um texto italiano relativamente simples com estrutura narrativa;
- 2) uma tradução do texto italiano para o francês;
- 3) um texto expositivo em francês sobre a geografia da França.

8.3.2 Amostras Aleatórias

Nas cinco amostras aleatórias descritas a seguir, sempre tratou-se de cursos de línguas na Universidade de Bochum no Ruhr:

- 1) O texto em italiano foi ministrado a 14 aprendizes de italiano relativamente adiantados.
- 2) a tradução francesa do texto italiano foi ministrada a 18 aprendizes de francês, com relativamente pouco conhecimento anterior.
- 3) o texto em francês, geograficamente orientado, foi apresentado a 23 estudantes com relativamente poucos conhecimentos anteriores (curso paralelo ao da amostra aleatória 2).
- 4) o texto A em francês, geograficamente orientado, ainda foi apresentado a 8 aprendizes relativamente adiantados.
- 5) Finalmente, o texto em francês, geograficamente orientado, foi ministrado a 8 participantes de um curso, o qual, ao contrário das pesquisas nas amostras aleatórias 1-4, visava exclusivamente a medição da competência local em aprendizes sem conhecimento anterior.

8.3.3 Método de Medição

A fim de medir a proporção do conteúdo da parte corretamente lembrado, nós subdividimos os textos nas chamadas unidades de conteúdo. Via de regra, estas unidades correspondiam a frases ou também a partes de frases. Acreditamos poder renunciar a uma análise proposicional formal, devido ao caráter exploratório de nossa pesquisa⁷. Tanto para o texto 1 como também para o texto 2 obtivemos 24 unidades de conteúdo.

No caso do texto 3, extraímos 17 unidades de conteúdo. Porém, infelizmente, verificou-se que uma coordenação entre unidades de conteúdo e segmentos de textos lembrados não foi possível, porque os sujeitos - diferentemente dos textos 1 e 2 - tendiam a gerar macro-posições neste texto expositivo. Por isso, avaliamos a medida de reprodução de texto totalmente em uma escala de 7 graus. A confiabilidade do inter-percentageal (medida com base na correlação de classe de Spearman de duas classes) foi de $r = .75$, no caso da amostra aleatória 4 e $r = .82$, no caso da amostra aleatória 5.

A avaliação das traduções ocorreu da seguinte maneira:

7 Isto também valeu rapidamente a métodos como "hierarchical content structure analysis" (conf. Connor, 1984), "topical structure analysis" (conf. Schneider & Connor, 1990), "story grammar approach" (conf. McCabe & Peterson, 1990; Walters & Wolf, 1986). Labovs "highpoint analysis" (conf. McCabe & Peterson, 1990) e medidas de complexidade lingüísticas (conf. Davison & Green, 1988).

Os sujeitos receberam 1 ponto para uma unidade de conteúdo - aqui designada como unidade de tradução - se a unidade, com limitações foi absoluta e corretamente traduzida. No caso de uma tradução errada, foi dado zero pontos e, meio ponto para uma tradução nem completamente errada, nem completamente certa.

Como outra variável importante e independente, ainda foi verificada a quantidade de aula que os sujeitos tiveram na respectiva língua, com base em um questionário ("conhecimento anterior" variável).

8.3.4 Análises Estatísticas e Resultados

Foram realizadas as seguintes análises:

- 1) Cálculo de valores médios e desvios padrão dos conhecimentos anteriores, dos resultados do Teste-C, bem como, dos resultados no "recall" e na tradução;
- 2) Cálculo das correlações entre os conhecimentos anteriores, os valores do Teste-C, os valores de "recall" e os valores de tradução;
- 3) Comparações múltiplas entre as médias dos conhecimentos anteriores, do resultado do Teste-C, do resultado do "recall" e do resultado da tradução;
- 4) Uma comparação individual dos sujeitos no que se refere a seus valores nas variáveis examinadas.

Quero, pelo menos, resumir o resultado principal de nossas pesquisas:

Com exceção da amostra aleatória 5, composta de sujeitos de um curso de leitura (e não - como nas restantes amostras al⁹⁹TVY3c cursos de línguas em geral), em todas as pesquisas houve uma correlação relativamente alta entre o número de itens corretamente resolvidos do Teste-C e a medida no conteúdo do texto corretamente lembrado. As respectivas correlações de classe de Spearman, bem como, as probabilidades unilaterais de excesso, sempre indicadas entre parênteses, são:

Amostra aleatória 1: .85 ($p < .001$);

Amostra aleatória 2: .71 ($p < .001$);

Amostra aleatória 3: .43 ($p = .026$);

Amostra aleatória 4: .82 ($p = .007$);

Amostra aleatória 5: -.30 ($p = .242$).

A correlação negativa de -.30 entre "recall" e as soluções do Teste-C na amostra aleatória 5, à primeira vista, parece um resultado bastante estranho. Apresentam-se pelo menos duas possibilidades de explicação:

- 1) O valor obtido não é significativo (a respectiva probabilidade de excesso é de 24%). Com isso, não podemos rejeitar a hipótese nula de uma correlação falha. De fato, o número de sujeitos com $n = 8$ é

muito pequeno nesta pesquisa e, com isso, a força do teste de significado não é muito alta. Por isso, em uma amostra aleatória maior poderia resultar um valor mais significativo.

- 2) Admitindo-se que não há uma correlação negativa entre o Teste-C e o "recall" na pesquisa 5, é necessário esclarecer, todavia, em relação às pesquisas restantes, por que não há nenhuma correlação positiva. Oferece-se o seguinte esclarecimento: os sujeitos na pesquisa 5 são estudantes de um curso de língua, que tem como finalidade exclusiva a mediação da competência de leitura e não a mediação de habilidades produtivas. Isto significa que, entre outros, estratégias de leitura também foram explicitamente passíveis de mediação. Como conseqüência, os estudantes podem conseguir um valor em pontos muito baixo no Teste-C, provavelmente, com base em habilidades produtivas falhas, especialmente na área da sintaxe e da morfologia, mesmo se eles entenderam relativamente bem o conteúdo do texto.

8.3.5 Conclusões Gerais

Acredito que, com base nas pesquisas realizadas por Grotjahn & Tönshoff (no prelo), os seguintes enunciados parecem justificados:

O Teste-C como método integrador, também carece de competência de leitura em uma determinada área. Por motivos plausíveis, isto realmente não vale para sujeitos como na amostra aleatória 5, ou seja, para aprendizes que não tiveram aula com utilização produtiva da língua. De fato, os resultados obtidos não justificam considerar o Teste-C como um teste de compreensão de leitura.

Os seguintes argumentos teóricos também se opõem ainda contra o Teste-C como compreensão de leitura:

- 1) Textos de Teste-C normalmente são muito curtos. Em conseqüência à capacidade muito importante para o processo normal de leitura, de construir macro-estruturas baseadas em textos mais longos, não podem ser medidas adequadamente com o auxílio do Teste-C.
- 2) Com base no formato do teste, a importante capacidade de passar rapidamente pelo texto, a fim de verificar exclusivamente o conteúdo global, para uma leitura com captação do sentido, não pode ser medida adequadamente com o Teste-C. Devido ao seu alto número de lacunas, o Teste-C é inferior ao Procedimento Cloze neste ponto.

Em suma, em vista de nossos resultados empíricos e dos argumentos teóricos apresentados, na minha opinião, deve-se partir do fato de que o Teste-C - especialmente em sua forma usual com apenas 20 à 25 lacunas e textos respectivamente curtos - não é um instrumento válido para a medição

da compreensão da leitura. Por isso, com base nos resultados até agora obtidos, também considero - ao contrário de Faust (1986) - que não se justifica o emprego do Teste-C como testes de compreensão de leitura.

8.4 As Pesquisas de Grotjahn (no prelo)

8.4.1 Princípios e Hipóteses

Geist (1987) sugere, que a diferença lingüística do texto normal entre coesão e coerência, deveria ser abandonada em favor do conceito de conectividade. Geis distingue três planos de conectividade em um texto: 1) o plano gramatical; 2) o plano temático e 3) o plano de discurso e intencionalidade. Como foi demonstrado através de inúmeras pesquisas empíricas, a elaboração ideal de um texto requer que este também esteja estruturado idealmente nos diferentes planos de conectividade. Correspondentemente, desvios de estrutura ótima podem levar a problemas de elaboração (conf., p. ex., as pesquisas de Darnell, 1963 e de Walters & Wolf, 1986). McCokie (1978, p.17) descreve este fato da seguinte maneira:

"... textual manipulations which reduce information useful to the reader in building a coherent representation of the content can be expected to reduce comprehension of the passage."

Uma argumentação análoga encontra-se em Halliday e Hasan (1976, p. 28):

"The sentences of a text ... are related to each other both substantively and by cohesion; and it is a characteristic of a text that the sequence of the sentences cannot be disturbed without destroying or radically altering the meaning. A text has meaning as a text ... Within a text the meaning of each sentence depends on its environment, including its cohesive relations with other sentences."

Com base nesses argumentos, espera-se, que modificações da ordem dos elementos de frases podem levar a problemas de elaboração (Scrambling of clauses) de um texto do Teste-C e, conseqüentemente, a um aumento de itens isolados e testes em geral. Porém, existem indicações empíricas, de que os Testes-C medem, sobretudo, no plano do micro-contexto (conf., p. ex., Cohen, Segal & Weiss Bar-Siman-Tov, 1984, 1985; Stemmer, 1991; Stemmer, no prelo)⁸.

⁸ Com referência ao Procedimento Cloze, também existe uma variedade de resultados correspondentes (conf. as referências literárias em Jonz, 1990). Porter (1983), também referindo-se ao Procedimento Cloze, fala de "a flaw in a measure of general proficiency". Opondo-se a isto, Jonz (1990, p. 61) afirma: "The research reported here demonstrates that... the standard fixed-ratio cloze procedure has a high level of sensitivity to intersentential ties and lexical selections..."

Isto poderia ocorrer devido ao fato de que os textos do Teste-C são muito curtos e, que, por isso, estratégias de elaboração macro-contextuais somente podem ser empregadas condicionalmente.

Em vista destes resultados, que podem ser submetidos empiricamente a testes, a seguinte hipótese pode ser formulada: Um "Scrambling" de frases parciais de um texto do Teste-C de extensão normal, não leva a um sensível aumento da dificuldade do teste. Porém, em textos mais longos do Teste-C, com uma estrutura seqüencial clara, pode-se contar com um aumento da dificuldade do teste.

8.4.2 Disposição do Experimento

Para a verificação destas hipótese, através do "scrambling" das sentenças parciais de três textos do Teste-C, eu elaborei três versões diferentes dos textos do Teste-C. A versão inicial não alterada do texto foi designada como Versão A. Na versão B, a seqüência das frases parciais, que juntas formaram uma unidade de conteúdo, não foi alterada. Como conseqüência, a Versão B tem maior conectividade do que a Versão C, na qual as frases parciais foram alteradas o mais amplamente possível.

Nos textos utilizados, tratava-se das narrativas em francês e do texto expositivo francês de Grotjahn & Tönshoff (no prelo).

Estes dois textos tinham 56 ou 52 lacunas. O terceiro texto era, como o texto 2, um texto expositivo. Porém, ele somente apresentou 28 itens e, com isso, teve quase o restante em lacunas. Todos os três textos mostraram uma ordem seqüencial clara em sua forma original.

Com o auxílio das versões do texto A, B e C, foram elaborados os seguintes três tipos de Testes-C: o Tipo 1 compunha-se dos textos 1A, 2B, 3C; o Tipo 2, dos textos 1B, 2C, 3A; e o Tipo 3, dos textos 1C, 2A, 3B. Cada tipo de teste, por conseguinte, continha um texto inalterado, um levemente alterado e um bastante alterado. Devido à sólida seqüência dos textos 1, 2 e 3, efeitos de posição potenciais estavam excluídos.

Os três tipos de testes foram apresentados a um total de 110 universitários(as) da Universidade de Bochum no Ruhr para estudo. A entrega dos três tipos de teste (condições) aos sujeitos foi feita ao acaso. A maioria dos sujeitos estudou francês no primeiro semestre e, em média, já havia tido aula de francês durante 4 anos.

8.4.3 Análises Estatísticas e Resultados

As análises estatísticas não forneceram nenhuma diferença entre os valores médios das versões de texto A, B e C e, portanto, nenhum efeito "scrambling". Porém, as análises demonstraram, simultaneamente, que os textos eram muito fáceis para os aprendizes relativamente adiantados. Visto que isto poderia ser um motivo para que o efeito "scrambling" não se

manifestasse, eu dividi a amostra aleatória total, com base na mediana do escore total dos três textos, em um grupo com maior capacidade e um, com menor capacidade.

Novamente, as análises estatísticas não forneceram nenhum efeito no grupo de maior capacidade. No grupo de menor capacidade, o "scrambling", entretanto, levou a um aumento da dificuldade do texto, como dito anteriormente no caso dos textos mais longos 1 e 2. No caso do terceiro texto, menor, não se pôde verificar nenhum efeito.

Análises de variação univariantes sempre com 2 e 52 graus de liberdade, resultaram em $F = 3.55$ e $p = .04$ para o texto 1, $F = 2.70$ e $p = .08$ para o texto 2 e $F = .29$ e $p = .75$ para o texto 3 (confira também as análises estatísticas mais profundas em Grotjahn, no prelo). Ao calcular a medida do efeito Omega2 (conf. Hays, 1973, p. 485), obtém-se .085 para o texto 1 e .058 para o texto 2. Isto significa, que nos textos 1 e 2, somente aproximadamente 8.5% ou 5.0% da variação da capacidade no Teste-C podem ser atribuídos às diferenças na conectividade do texto. Um possível efeito "scrambling", por conseguinte, também está muito vagamente definido no caso dos textos 1 e 2.

Em suma, as pesquisas de Grotjahn (no prelo) parecem confirmar a hipótese de que os Testes-C medem proporções normais, sobretudo, no plano do micro-contexto. Ao mesmo tempo, os resultados indicam, que os processos de elaboração macro-contextuais centrais para uma leitura com captação do sentido, possivelmente, representam um papel mais forte em textos mais longos de teste-C. Isto significaria que, com ajuda de textos mais longos, um aspecto muito importante do processo de leitura (língua estrangeira) pode ser testado. Na realidade, para um maior esclarecimento desta questão, inúmeras pesquisas com textos de diferentes extensões e de estrutura diferente ainda são necessários. Nesta relação, também poderia ser examinado se os Testes-C, nos quais somente as palavras de conteúdo e/ou as palavras de função centrais foram prejudicadas, apresentam métodos válidos para a medição da competência de leitura.

9. Conclusão

O estado atual da pesquisa em relação ao Teste-C pode ser resumido da seguinte maneira: o Teste-C é um instrumento de medição relativamente econômico, muito objetivo e principalmente muito confiável. A capacidade básica que o Teste-C mede com precisão, ou seja, sua validade de construto, até agora ainda não foi suficientemente esclarecida. Esta crítica também vale em relação a outros testes de língua indiretos ou semi-indiretos.

Na prática, o Teste-C teria condições de ser implantado com êxito, principalmente, se o objetivo fosse um diagnóstico global da situação da língua, amplamente independente da instrução progressiva, sobretudo, ligado às habilidades da linguagem literária. Em comparação com os testes discrete-point padronizados, o Teste-C tem a vantagem de que ele é relativamente fácil de desenvolver. Em comparação com entrevistas, ele tem a vantagem de que ele pode ser inserido como teste de grupos e, que há uma alta objetividade de execução e de avaliação. Além disso, ao contrário da entrevista, nem a execução, nem a avaliação pressupõe competência em língua estrangeira com variantes aceitáveis alegadas - pelo menos no caso de exact scoring ou acceptable scoring. Isto é uma vantagem que não se pode desprezar, sobretudo, no caso de uma aula com professores que não falam a língua materna. Além disso, a execução do Teste-C não requer meios técnicos como retroprojetor ou áudio. Como todos sabem, no que diz respeito à execução de testes, esta é uma vantagem que não pode ser desprezada.

Em suma, finalizando, pode-se verificar que o Teste-C, apesar de certas limitações, representa o enriquecimento extremamente fecundo no panorama de testes disponíveis.

REFERÊNCIAS BIBLIOGRÁFICAS

- Afferbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior*, 16, 307-322.
- Alderson, J. Charles (1978). *A study of cloze procedure with native and non-native speakers of English*. Unpublished Ph. D. dissertation. Edinburgh: University of Edinburgh.
- Alderson, J. Charles (1979). The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13, 219-227.
- Alderson, J. Charles, & Urganhart, A.H. (eds.) (1984). *Reading in a foreign language*. London: Longman.
- Bachman, Lyle F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Ballstaedt, Steffen-Peter, Mandl, Heinz, Wolfgang, & Tergan, Sigmar-Olaf (1981). *Texte verstehen, Texte gestalten*. München: Urban & Schwarzenberg.
- Baur, Rupprecht S. Meder, Gregor (1989). Die Rolle der Muttersprache bei der schulischen Sozialisation ausländischer Kinder. *Diskussion Deutsch*, 20, 119-135.
- Baur, Rupprecht S. & Meder, Gregor (in press). C-Tests zur Ermittlung der globalen sprachfähigkeit im Deutschen und in der Muttersprache bei ausländischen Schülern in der BRD. In Rüdiger Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Bernhardt, Elizabeth B. (1986). Reading in the foreign language. In Barbara H. Wing (Ed.), *Listening reading writing: Analysis and application* (pp. 93-115). Middleburg, VT: Northeast Conference on the Teaching of Foreign Languages.
- Blook, Ellen (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463-494.
- Bolten, Jürgen (in press). Wie schwierig ist ein C-Test? Erfahrungen mit dem C-Test als Einstufungstest in Hochschulkursen Deutsch als Fremdsprache. In Rüdiger Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.

- Chapelle, Carol A., & Abraham, Roberta G. (1990). Cloze method: what difference does it make? *Language Testing*, 7, 131-146.
- Clary, Christopher (1988). The C-Test in English: left-hand deletions. *RELC Journal*, 19, 26-37.
- Cohen, Andrew D. (1984a). On taking language tests: what the students report. *Language Testing*, 1, 10-81.
- Cohen, Andrew D. (1984b). The use of mentalistic measures in determining LSP reading problems. In A.K. Pugh & Jan M. Ulijn (Eds.), *Reading for professional purposes. Studies and practices in native and foreign languages* (pp. 177-189). London: Heinemann Educational Books.
- Cohen, Andrew D. (1986). Mentalistic measures in reading strategy research: Some recent findings. *English for Specific Purposes*, 5, 131-145.
- Cohen, Andrew D. (1987). Recent uses of mentalistic data in reading strategy research. *Documentação de Estudos em Lingüística Teórica e Aplicada (D.E.L.T.A.)*, 3, 57-84, (Dept. of Linguistics, Pontifical Catholic University of São Paulo, Brasil).
- Cohen, Andrew D., Segal, Michal, & Weiss Bar-Siman-Tov, Ronit (1984). The C-Test in Hebrew. *Language Testing*, 1, 221-225.
- Cohen, Andrew D., Segal, Michal, & Weiss Bar-Siman-Tov, Ronit (1985). The C-Test in Hebrew. *Fremdsprachen und Hochschule*, 13/14, 121-127.
- Connor, Ulla (1984). Recall of Text: Differences between first and second language readers. *TESOL Quarterly*, 18, 239-250.
- Commins, Jim (1983). Language proficiency and academic achievement. In John W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 108-129). Rowley, Mass.: Newbury House.
- Darnell, D.K. (1963). The relation between sentence order and comprehension. *Speech Monographs*, 30, 97-100.
- Davison, Alice, & Green, Georgia M. (eds) (1988). *Linguistic complexity and text comprehension: Readability issues reconsidered*. Hillsdale, NJ: Lawrence Erlbaum.
- Ericsson, K. Anders (1988). Concurrent verbal reports on text comprehension: A review. *Text*, 8, 295-325.
- Farch, Claus, & Kasper, Gabriele (Eds.) (1987). *Introspection in second language research*. Clevedon: Multilingual Matters.
- Faust, Berthold (1986). Der C-Test: Eine Methode zur Messung der Sprachbeherrschung. *Spracharbeit*, 3, 10-17.
- Feldmann, Ute; Grotjahn, Rüdiger, & Stemmer, Brigitte (1986). Was massen Sprachtests eigentlich? Überlegungen zur introspektiven Validierung von C-Tests. In Seminar für Sprachlehrforschung der Ruhr-Universität Bochum (Ed.), *Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre* (pp. 325-338). Frankfurt/M.: Scriptor.
- Feldmann, Ute & Stemmer, Brigitte (1987). Thin-alud a-retrospective da- in C-te- taking: diffe-languages- diffe- learners - sa- approaches? In Klaus Faerch & Gabriele Kasper (Eds.), *Introspection in second language research* (pp. 251-267). Clevedon: Multilingual Matters.
- Gaies, Stephen J. (1987). Validation of the noise test. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests*. (pp. 41-74). Bochum: Brockmeyer.
- Garner, Ruth (1987). *Metacognition and reading comprehension*. Norwood, NJ: Ablex.
- Geist, Uwe (1987). The three levels of connectivity in a text. *Journal of Pragmatics*, 11, 737-749.
- Graesser, Arthur C. & Bower, Gordon H. (Eds.) (1990). Inferences Grönwoldt, Peter (1985). Erfahrungen mit dem C-Test am Eurozentrum Köln. *Fremdsprachen und Hochschule*, 13/14, 91-95.
- Grotjahn, Rüdiger (1986). Test validation and cognitive psychology: some methodological considerations. *Language Testing*, 3, 180-185.
- Grotjahn, Rüdiger (1987a). How to construct and evaluate a C-Test: A discussion of some problems and some statistical analyses. In Rüdiger Grotjahn, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Taking their measure: The validity and validation of language tests* (pp. 219-253). Bochum: Brockmeyer.
- Grotjahn, Rüdiger (1987b). Ist der C-Test ein Lasetest? In Anthony Addison & Klaus Vogel (Eds.), *Lehren und Lernen von Fremdsprachen in Studium* (pp. 230-248), Ruhr-Universität Bochum: AKS.
- Grotjahn, Rüdiger (1987c). On the methodological basis of introspective methods. In Claus Faerch & Gabriele Kasper (Eds.), *Introspection in second language research* (pp. 54-81). Clevedon: Multilingual Matters.
- Grotjahn, Rüdiger (1969). Der C-Test im Bundeswettbewerb Fremdsprachen: Eignung und Probleme (Französisch). In Thomas Finkenstaedt & Konrad Schröder (Eds.), *Zwischen Empirie und Machbarkeit: Erstes Symposium zum Bundeswettbewerb Fremdsprachen* (pp. 41-56). Augsburg: Universität.
- Grotjahn, Rüdiger (in press). 'Scrambled' C-Tests: Untersuchungen zum Zusammenhang zwischen Lösungsgüte und sequentieller Textstruktur. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Grotjahn, Rüdiger; Hohenbleicher, Heike & Tönshoff, Wolfgang (in press). Der C-Test im Italienischen: Theoretische Überlegungen und empirische Analysen. In Rüdiger Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Grotjahn, Rüdiger & Stemmer, Brigitte (1985). On the development and evaluation of a C-Test for French. *Fremdsprachen und Hochschule*, 13/14, 101-120).
- Grotjahn, Rüdiger & Tönshoff, Wolfgang (in press). Textverständnis bei der C-Test-Bearbeitung. Pilotstudien mit Französisch und Italienischlernern. In Rüdiger Grotjahn (Ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Halliday, Michael A.K. & Hasan, Rukaiya (1976). *Cohesion in English*. London: Longman.
- Hays, William L. (1973). *Statistics for the social sciences*. London & New York: Holt, Rinehart and Winston.
- Herrmann, Theo, & Hoppe-Graff, Siegfried (1988). Textproduktion. In Heinz Mands & Hans Spada (Eds.), *Wissenspsychologie* (pp. 293-298). München: Psychologie Verlagsunion.
- Hosenfeld, Carol (1984). Case studies of ninth grade readers. In Klein-Braley, Christine (1985b). C-Tests and construct validity. *Fremdsprachen und Hochschule*, 13/14, 55-65.
- Klein-Braley, Christine (1985c). C-Tests as placement tests for German university students of English. *Fremdsprachen und Hochschule*, 13/14, 96-100.
- Klein-Braley, Christine (1985d). Reduce redundancy as an approach to language testing. *Fremdsprachen und Hochschule*, 14/14, 1-13.
- Klein-Braley, Christine, (1985e). Tests of reduced redundancy - theory. In viljo kohonen & Antti J. Pitkänen (Eds.), *Language testing in shool. APinLA Yearbook 1985* (pp. 33-48). Tampere: AFinLA.
- Klein-Braley, Christine & Raatz, Ulrich (1982). Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung. *AKS-Rundbrief*, 4, 23-37.
- Klein-Braley, Christine, & Raatz, Ulrich (1984). A survey of research on the C-Test. *Language Testing*, 1, 134-146.
- Klein-Braley, Christine & Raatz, Ulrich (Eds.), (1985), C-Test in der Praxis (thematic volume). *Fremdsprachen und Hochschule*, 13/14. Ruhr-Universität Bochum; AKS.
- Klume, Rainer H. (1988). Methoden der Psychologie zur Gewinnung von Daten über menschliches Wissen. In Heins Mandl & Hans Spada (Eds.), *Wissenspsychologie* (pp. 359-385). München: Psychologie Verlagsunion.
- foreign language* (pp. 231-244). London: Longman.
- Jakschik, Gerhard (in press). Zum Einsatz des C-Test in den Psychologischen Diensten der Arbeitsämter: Ein C-Test fürDeutsch als Zweitsprache. In Rüdiger Grotjahn (Ed.), *Der C-test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Johns, Ann A. (1985). Summary protocols of "underprepared" and "adept" university students: replications and distortions of the original. *Language Learning*, 36, 495-517.
- Johnston, Peter H. (1984). Assessment in reading. In David P. Pearson (ed.), *Handbook of reading research* (pp. 147-182). New York: Longman.
- Jonz, John (1990). Another turn in the conversation: What does cloze measure? *TESOL Quarterly*, 24, 61-83).

- Klein-Braley, Christine (1981). *Empirical investigations of cloze tests: an examination of the validity of cloze tests as tests of general language proficiency in English for German university students*. Unpublished Ph. D. dissertation. Universität Duisburg.
- Klein-Braley, Cristine (1982). On the suitability of cloze tests as measures of reading comprehension. *Toegepaste Taalwetenschap in*, 13, 49-81.
- Klein-Braley, Christine (1985a). A cloze-up on the C-test: A study in the construct validation of authentic tests. *Language*
- Kokkoto, Valmar (1988). Letter deletion procedure: a flexible way of reducing text redundancy. *Language Testing*, 5, 115-119.
- Langer, Judith A. (1987). The construction of meaning and the assessment of comprehension: An analysis of reader performance on standardized test items. In Roy O. Freedle & Richard P. Duran (Eds.), *Cognitive and linguistic analyses of test performance*. (pp. 225-244). Norwood, NJ: Ablex.
- Lee, James F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 201-211.
- Licnert, Gustav A. (1969). *Testaufbau und Testanalyse* (3rd. ed.) Weinheim: Beltz.
- McCabe, Alyssa & Peterson, Carola (1990). What makes a narrative memorable? *Applied Psycholinguistics*, 11, 73-82.
- MaConkie, George W. (1978). Learning from text. *Review of Research in Education*, 5, 3-48.
- Moosbrugger, Helfrid & Müller, Hans (1981). Ein Klassisches Latent-additives Testmodell (KLA). In Wolfgang Michaelis (Ed.), *Bericht über den 32. Kongress der Deutschen Gesellschaft für Psychologie in Zürich 1980* (Vol. 2, pp. 383-486). Göttingen, Toronto, Zürich: Hogrefe.
- Moosbrugger, Helfrid & Müller, Hans (1982). A classical latent additive test model (CLA model). *The German Journal of Psychology*, 6, 145-149.
- Oller, John W. Jr. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23, 105-118.
- Oller, John W. Jr. (1976). Evidence for a general language proficiency factor: and expectancy grammar. *Die Neueren Sprachen*, 75, 165-274.
- Oller, John W. Jr. (1979). *Language tests at school*. London: Longman.
- Olson, Gary M.; Muffy, Susan A. & Mack, Robert L. (1984). Thinking-out-loud as a method for studying real time comprehension processes. In David F. Kieras & Marcel A. Just (Eds.), *New methods in reading comprehension research* (pp. 253-285). Hillsdale, N.J.: Lawrence Erlbaum.
- Phéh, Csaba; Oláh, Attila & Zétényi, Tamás (1988). Individual differences in story recall. *Tex*, 8, 395-410.
- Porter, Don (1983). The effect of quantity of context on the ability to make linguistic-predictions: a flaw in a measure for general proficiency. In Arthur Hughes & Don Porter (Eds.), *Current development in language testing* (pp. 63-74). London: Academic Press.
- Raatz, Ulrich (1982). Language theory and factor analysis. In Madeline Lutjeharms & Terry Culhane (Eds.), *Practice and problems in language testing 3* (pp. 30-56). Brussel: Vrije Universiteit.
- Raatz, Ulrich (1984). The factorial validity of C-Test. In Terry Culhane, Christine Klein-Braley & Douglas K Stevenson (Eds.), *Practice and problems in language testing 7* (pp. 124-139). Colchester: University of Essex, Dept. of Language and Linguistics.
- Raatz, Ulrich (1985a). Better theory for better tests? *Language Testing*, 2, 60-75.
- Raatz, Ulrich (1985b). Investigating dimensionality of language tests - a new solution to an old problem. In Yijio Kohonen, Kilikka von Essen & Christine Klein-Braley (Eds.), *Practice and problems in language testing 8* (pp. 123-136). Tampere: FFinLA.
- Raatz, Ulrich (1985c). Tests of reduced redundancy the C-Test, a practical example. In Viljo Kohonen & Antti J. Patkainen (Eds.), *Language testing in school. AFinLA Yearbook 1985* (pp. 49-62). Tampere: AFinLA.
- Raatz, Ulrich (1985d). The factorial validity of C-Tests. *Fremdsprachen und Hochschule*, 13/14, 42-54.
- Raatz, Ulrich & Klein-Braley, Christine (1982). The C-test - a modification of the cloze procedure. In Terry Culhane, Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Practice and problems in language testing IV* (pp. 113-138). Colchester: University of Essex, Dept. of Language and....
- Raatz, Ulrich & Klein-Braley, Christine (1989). Ein zur Messung der Sprachleistung. Der C-Test: Theorie und Praxis. In Ralf Horn, Karlheinz Ingenkamp & Reinhold S. Jäger (Eds.), *Tests und Thends 3. Jhrbuch der Pädagogischen Diagnostik* (pp. 107-138). Weinheim & Basel: Beltz.
- Raatz, Ulrich, & Klein-Braley, Christine (1985). How to develop a C-Test. *Fremdsprachen und Hochschule*, 13/14, 20-22.
- Raatz, Ulrich, & Klein-Braley, Christine (1989). Der C-Test im Bundeswettbewerb Fremdsprachen: Eingnung und Probleme (Englisch). In Thomas Finkenstaedt & Konrad Schröder (Eds.), *zwischen Empirie aund Machbarkeit: Erstes Symposium zum Bundeswettbewerb Fremdsprachen* (pp. 127-134). Augsburg: Universität.
- Rickheit, Gert & Strohner, Hans (1989). Textreproduktion. In Gerd Antos & Hans Krigs (Eds.), *Textproduktion. Ein interdisziplinärer Forschungsüberblick* (pp. 222-256). Tübingen: Niemeyer.
- Rickheit, Gert & Strohner, Hans (1990). Inferenzen: Basis des Sprachverstehens. *Die Neueren Sprachen*, 89, 532-545.
- Rickheit, Gert & Strohner, Hans (Eds.) (1985). *Inferences in text processing*. Amsterdam: North Holland.
- Roos, Undine (in Press). Einige Probleme bei der Anwendung des C-Test-Prinzips auf das Japanische. In Rüdiger Grotjahn (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Schneider, Melanie, & Connor, Ulla (1990). Analyzing topical structure in ESL essays: not all topics are equal. *Studies in Second Language Aquisition*, 12, 411-427.
- Spolsky, Bernard (1973). What does it mean to know a language or how do you get someone to perform his competence? In John W. Oller, Jr. & Jack C. Richards (Eds.), *Focus on the learner* (pp. 164-176). Rowley, Mass.: Newbury House.
- Steffensen, Margaret S., & Joang-Dev, Chitra (1984). Cultural Knowledge and reading. In J. Charlot Alderson & A.H. Urguhart (Eds.), *Reading in a foreign language* (pp. 48-61). London: Longman.
- Stemmer, Brigitte (1991). *What's on a C-test taker's mind: Mental processes in C-test taking*. Bochum: Brockmeyer.
- Stemmer, Brigitte (in press). An alternative approach to C-test validation. In Rüdiger Grotjahn (Ed.) *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: Brockmeyer.
- Sternglass, Marilyn S. (1988). The presence of thought. *Introspective accounts of reading and writing*. Norwood, NJ: Ablex.
- Tuinman, J. Jaap (1973/1974). Determining the passage-dependency of comprehension questions in 5 major test. *Reading Research Quarterly*, 9(2), 206-223.
- Vollmer, Helmut. (1981). Why are we interested in 'General Language Proficiency'? In Christine Klein-Braley & Douglas K. Stevenson (Eds.), *Practice and problems in language testing 1* (pp. 96-123). Bern & Frankfurt/M.: Lang.
- Vollmer, Helmut J. (1983). The structure of foreign language competence. In Arthur Hughes & Don Porter (Eds.), *Current developments in language testing* (pp. 3-29). London: Academic Press.
- Vollmer, Helmut J., & Sang, Fritz (1980). Zum psycholinguistischen Konstrukt einer internalisierten Erwartungsgrammatik. *Linguistik und Didaktik*, 42, 122-148.
- Walters, Joel, & Wolf, Yuval (1986). Language proficiency, text content and order effects in narrative recall. *Language Learning*, 36, 47-64.
- Weaver, Wendell; Bickley, Albert C. & Ford, Fraughton (1969). A cross-validation study of the relationship of reading test items to their relevant paragraphs. *Perceptual and motor skills*, 20, 11-14.
- Wolff, Dieter (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition*, 9, 307-326.

PUBLICAÇÕES PERIÓDICAS DA PUCRS

VERITAS

Revista de cultura geral – Trimestral

LETRAS DE HOJE

Revista de estudos de Lingüística, Literatura e Língua Portuguesa – Trimestral

TEOCOMUNICAÇÃO

Revista de estudos de Teologia, Filosofia e áreas afins, órgão de comunicação do Instituto de Teologia – Trimestral

ESTUDOS IBERO-AMERICANOS

Revista de estudos sobre a História e a Literatura Ibero-Americana, do Curso de Pós-Graduação em História – Semestral

REVISTA DE MEDICINA DA PUCRS

Editada pela Faculdade de Medicina e Instituto de Geriatria – Trimestral

PSICO

Revista especializada em Psicologia – Semestral

DIREITO & JUSTIÇA

Revista da Faculdade de Direito – Sem periodicidade

EDUCAÇÃO

Revista do Curso de Pós-Graduação em Educação – Semestral

ODONTO CIÊNCIA

Revista da Faculdade de Odontologia – Semestral

PUCRS – INFORMAÇÃO

Boletim informativo – Bimestral

AGENDA PUCRS

Boletim informativo interno da PUCRS – Mensal

COMUNICAÇÕES DO MUSEU DE CIÊNCIAS

Sem periodicidade

MUNDO JOVEM

Jornal de idéias e reflexões para jovens, vinculado ao Instituto de Teologia e Ciências Religiosas – Mensal

ANÁLISE

Revista da Faculdade de Ciências Políticas e Econômicas – Semestral

