

PENGEMBANGAN TES SUMATIF BAHASA JERMAN SEMESTER GANJIL KELAS X SMA NEGERI 1 TAMAN

Mas Uliatul Hikmah

Mahasiswa Program Studi S1 Pendidikan Bahasa Jerman, Fakultas Bahasa dan Seni,
Universitas Negeri Surabaya

mashikmah16020094034@mhs.unesa.ac.id

Drs. H. Benny Herawanto Soesetyo, M.Psi.

Dosen Program Studi S1 Pendidikan Bahasa Jerman, Fakultas Bahasa dan Seni, Universitas
Negeri Surabaya

ABSTRAK

Evaluasi merupakan proses pengumpulan data untuk memberikan gambaran perkembangan belajar peserta didik. Untuk mencapai tujuan dalam evaluasi, maka diperlukan suatu instrumen yang tepat karena instrumen evaluasi yang kurang baik akan mengakibatkan hasil penilaian menjadi tidak sesuai dengan realita. Evaluasi pembelajaran bahasa Jerman di SMA Negeri 1 Taman salah satunya adalah tes sumatif yang dilaksanakan tiap akhir semester. Dalam mengembangkan tes yang baik dan dapat dipertanggungjawabkan, maka perlu dianalisis validitas, reliabilitas, tingkat kesukaran, daya pembeda dan efektivitas distraktor (Nurgiyantoro, 2010). Sehingga rumusan masalah dalam penelitian ini adalah bagaimana bentuk tes sumatif bahasa Jerman kelas X SMA Negeri 1 Taman yang sesuai kriteria. Tujuan dari penelitian ini adalah menghasilkan tes sumatif bahasa Jerman kelas X SMA Negeri 1 Taman yang sesuai kriteria. Kriteria yang telah ditetapkan adalah a) butir soal yang dikembangkan valid, b) butir soal yang dikembangkan reliabel, c) tingkat kesukaran kategori sedang d) daya pembeda butir soal minimal cukup, dan e) distraktor pada soal efektif. Dari kriteria tersebut, butir soal dapat diterima apabila memenuhi minimal tiga kriteria termasuk validitas dan reliabilitas. Butir soal ditolak apabila tidak memenuhi semua kriteria. Untuk mengembangkan instrumen evaluasi perlu melalui serangkaian langkah-langkah pengembangan tes sumatif yang telah ditetapkan. Sehingga dapat diperoleh instrumen evaluasi tes sumatif yang dapat dipertanggungjawabkan kualitasnya. Penelitian ini menggunakan penelitian kuantitatif dengan jenis penelitian uji coba. Populasi dalam penelitian ini adalah kelas peserta didik XI SMA Negeri 1 taman yang menempuh mata pelajaran bahasa jerman sebagai lintas minat maupun pelajaran wajib. Sedangkan sampel dalam penelitian ini adalah peserta didik kelas XI Bahasa SMA Negeri 1 Taman. Hasil dari penelitian ini adalah kisi-kisi tes sumatif pilihan ganda berdasarkan kurikulum 2013 dan taksonomi bloom revisi aspek kognitif dan aspek psikomotor. Dari 50 butir soal yang telah dikembangkan, analisis kuantitatif tes tersebut menunjukkan sebanyak 20 butir tes dapat diterima karena memenuhi kriteria yaitu nomor 1, 3, 4, 6, 11, 12, 14, 16, 17, 19, 21,23, 24, 31, 32, 33, 40, 44, 47, dan 48. Dan sebanyak 30 butir soal ditolak karena tidak memenuhi kriteria yaitu nomor 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50.

Kata Kunci : Pengembangan Tes, Tes Sumatif Bahasa Jerman, Semester Ganjil

ABSTRACT

Evaluation is a process of collecting facts to give an overview of students' learning progress. To achieve the goals of the evaluation, an appropriate instrument is needed because poor evaluation instruments will result in the assessment results being not by reality. Evaluation of German learning in SMA Negeri 1 Taman one of which is a summative test that is carried out at the end of each semester. In developing a good and accountable test, it is necessary to analyze the validity, reliability, level of difficulty, differentiation and effectiveness of the distractor (Nurgiyantoro, 2010). So that the formulation of the problem in this study is how the form of summative tests of German language class X SMA Negeri 1 Taman that fits the criteria. The purpose of this study is to produce a summative test of German language class X SMA Negeri 1 Taman that fits the criteria. The predetermined criteria are a) the items developed are valid, b) the items developed are reliable, c) the level of difficulty is moderate c) the distinguishing power of items is minimally sufficient, and e) the distractor is effective. From these criteria, items can be accepted if they meet a minimum of three criteria including validity and reliability. The items are rejected if they do not meet all the criteria. To develop an evaluation instrument it is necessary to go through a series of steps in the development of a predetermined summative test. So that it can be obtained summative test evaluation instruments that can be accounted for quality. This research uses quantitative research with the type of trial research. The population in this study was class XI SMA Negeri 1 Taman, which took German language as a cross-interest and compulsory subject. While the sample in this study was class XI SMA Negeri 1 Taman. The results of this study are a multiple-choice summative test grid based on the 2013 curriculum and bloom revised taxonomy of cognitive and psychomotor aspects. Of the 50 items that have been developed, the quantitative analysis of the test shows as many as 20 test items can be accepted because it meets the criteria of numbers 1, 3, 4, 6, 11, 12, 14, 16, 17, 19, 21,23, 24, 31, 32, 33, 40, 44, 47, and 48. And as many as 30 items were rejected because they did not meet the criteria of number 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50.

Keywords : Test Development, German Summative Test, Odd Semester

PENDAHULUAN

Evaluasi merupakan proses pengumpulan data untuk memberikan gambaran perkembangan belajar peserta didik. Gambaran perkembangan tersebut perlu diketahui oleh pendidik untuk memastikan apakah peserta didik telah mengalami proses pembelajaran dengan benar atau tidak. Kemudian pendidik dapat menentukan alternatif yang tepat dalam pengambilan keputusan. Untuk mencapai tujuan dalam evaluasi, maka diperlukan suatu instrumen yang tepat karena instrumen evaluasi yang kurang baik akan mengakibatkan hasil penilaian menjadi tidak sesuai dengan realita. Bahasa Jerman di SMA Negeri 1 Taman merupakan mata pelajaran yang wajib ditempuh oleh peserta didik jurusan bahasa dan lintas minat wajib. Untuk mengetahui gambaran perkembangan belajar peserta didik, pendidik melakukan evaluasi berbentuk tes tes sumatif yang dilaksanakan tiap akhir semester dengan bentuk tes pilihan ganda. Menurut Nurgiyantoro (2010, 26), dalam mengembangkan tes yang baik dan dapat dipertanggungjawabkan, maka tes dianalisis berdasarkan validitas dan reliabilitasnya serta tingkat kesukaran, daya pembeda dan efektivitas distraktor. Dari hasil analisis tersebut akan diperoleh informasi empirik yang menunjukkan kualitas tes. Berdasarkan persyaratan tes sumatif di atas, pendidik menginginkan adanya sebuah tes sumatif yang tepat atau sesuai kriteria. Maka penelitian ini merupakan penelitian pengembangan tes sumatif yang ditujukan khusus untuk kelas X semester ganjil sebagai pemula dalam mempelajari bahasa Jerman.

Rumusan masalah dalam penelitian ini yaitu bagaimana bentuk tes sumatif bahasa Jerman kelas X SMA Negeri 1 Taman yang sesuai kriteria. Sedangkan tujuan penelitian dalam penelitian ini yaitu menghasilkan tes sumatif bahasa Jerman kelas X SMA Negeri 1 Taman yang sesuai kriteria. Kriteria butir soal mencakup; a) butir soal yang dikembangkan valid, b) butir soal yang dikembangkan reliabel, c)

tingkat kesukaran kategori sedang, d) daya pembeda butir soal minimal cukup, dan e) distraktor pada soal efektif. Berdasarkan kriteria di atas, butir soal diterima atau ditolak apabila sebagai berikut; 1) butir soal diterima apabila memenuhi minimal tiga kriteria termasuk validitas dan reliabilitas, dan 2) butir soal ditolak apabila tidak memenuhi semua kriteria.

Batasan masalah untuk menghindari kesalahpahaman dalam penelitian ini adalah sebagai berikut;

1. Instrumen disajikan dalam bentuk soal tes sumatif berbentuk tes pilihan ganda yang dikembangkan berdasarkan silabus dan KD pada K13
2. Soal Tes dikembangkan dengan menetapkan kisi-kisi soal.
3. Pengembangan yang dimaksud dalam penelitian ini adalah pembuatan tes sumatif berdasarkan langkah-langkah pengembangan tes sumatif.
4. Pengolahan data yang digunakan menggunakan program pengolahan data Microsoft Excel 2010.
5. Tes Sumatif diujikan kepada peserta didik kelas XI SMA Negeri 1 Taman karena telah mendapatkan keseluruhan materi kelas X Semester Ganjil.

METODE

Penelitian ini menggunakan pendekatan penelitian kuantitatif. Pendekatan kuantitatif dilakukan untuk mendapatkan data empiris kualitas tes sumatif dari hasil pekerjaan peserta didik. Data empiris yang diperoleh berupa angka kemudian dianalisis menggunakan bantuan komputer dengan program *Microsoft Excel*. Sehingga jenis penelitian yang digunakan dalam penelitian ini adalah penelitian uji coba. Populasi dalam penelitian ini adalah peserta didik kelas XI SMA Negeri 1 Taman yang menempuh mata pelajaran bahasa Jerman sebagai lintas minat maupun pelajaran wajib. Sedangkan Sampel yang digunakan dalam penelitian ini adalah kelas XI Bahasa dengan pertimbangan bahasa Jerman merupakan

mata pelajaran wajib untuk jurusan Bahasa.

Teknik analisis data pada penelitian ini menggunakan analisis kuantitatif dengan mengukur beberapa variabel yaitu, validitas butir soal, reliabilitas tes, indeks tingkat kesukaran soal, indeks daya pembeda soal dan efektifitas distraktor. Dalam menganalisis data kuantitatif ini menggunakan aplikasi Microsoft Excel sesuai dengan rumus masing-masing variabel.

Validitas Butir Soal

Validitas adalah mengukur ketepatan apa yang seharusnya diukur. Validitas butir soal adalah mengukur ketepatan butir soal dengan aspek yang diukur. Cara memperoleh angka validitas butir soal adalah dengan menghitung indeks korelasi antara skor tiap butir dengan skor totalnya. Satu butir soal dikatakan valid apabila skor tersebut sesuai dengan skor totalnya. Sehingga ada korelasi positif antara skor butir soal dengan skor total. Validitas butir soal pilihan ganda dihitung menggunakan korelasi point biserial. Kemudian dikonsultasikan dengan r_{tabel} pada taraf signifikansi 5% sesuai dengan jumlah peserta didik yang diteliti. Butir soal dikatakan valid apabila $r_{pbi} > r_{tabel}$, jika r_{pbi} kurang dari atau sama dengan r_{tabel} , maka butir soal tersebut tidak valid. Validitas butir soal bentuk pilihan ganda dapat dihitung menggunakan rumus korelasi poin biserial:

$$r_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Keterangan:

- r_{pbi} = koefisien korelasi biserial
- M_p = rerata skor dari subjek yang menjawab betul bagi item yang dicari validitasnya
- M_t = rerata skor total
- S_t = standar deviasi dari skor total proporsasi
- p = proporsi peserta didik yang menjawab benar

$$p = \frac{\text{banyaknya siswa yang menjawab benar}}{\text{jumlah seluruh peserta didik}}$$

q = proporsi siswa yang menjawab salah ($q = 1 - p$)

(Arikunto, 2013)

Reliabilitas tes

Reliabilitas tes dilakukan untuk mengetahui tingkat keajekan skor tes atau tingkat konsistensi antar item soal. Reliabilitas soal dapat diketahui dengan melihat besar kecil koefisien reliabilitas soal tersebut. Reliabilitas soal pilihan ganda dapat diketahui dengan menggunakan rumus KR-20. Soal dikatakan reliable apabila koefisien reliabilitas soal (r_{11}) sama dengan atau lebih besar dari 0,70. Semakin tinggi koefisiennya maka semakin reliabel soal tersebut. Reliabilitas dapat dihitung menggunakan rumus KR-20 sebagai berikut:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \Sigma pq}{S^2} \right)$$

Keterangan:

- r_{11} = reliabilitas tes secara keseluruhan
- p = proporsi subjek yang menjawab item dengan benar
- q = proporsi subjek yang menjawab item dengan salah ($q = 1 - p$)
- Σpq = jumlah hasil perkalian antara p dan q
- n = banyaknya item
- S = standar deviasi dari tes

(Arikunto, 2013)

Tingkat Kesukaran

Tingkat kesukaran adalah peluang dalam menjawab benar suatu soal pada tingkat kemampuan tertentu yang dinyatakan dalam bentuk indeks. Indeks kesukaran dinyatakan dalam bentuk proporsi antara 0,00-1,00. Semakin kecil angka indeks kesukaran, maka semakin susah soal tersebut. Butir soal dapat dikatakan sebagai butir soal yang baik apabila tidak terlalu sukar dan tidak terlalu mudah. Jadi

butir soal yang baik memiliki kesukaran yang sedang atau cukup dengan indeks kesukaran antara 0,30-0,70. Angka indeks kesukaran dapat diperoleh dengan menggunakan rumus sebagai berikut:

$$P = \frac{B}{JS}$$

Keterangan:

P = angka indeks kesukaran item
B = banyaknya testee yang dapat menjawab dengan betul terhadap butir item yang bersangkutan
JS = jumlah testee yang mengikuti tes hasil belajar

Kriteria indeks kesukaran soal adalah sebagai berikut:

0,00-0,30 = soal kategori sukar

0,30-0,70 = soal kategori sedang

0,70-1,00 = soal kategori mudah

(Arikunto, 2013)

Daya Pembeda

Daya pembeda dapat menunjukkan perbedaan antara peserta didik yang telah menguasai materi dengan yang kurang. Daya pembeda soal dapat diketahui dengan melihat besar kecilnya angka indeks diskriminasi soal. Indeks diskriminasi soal berkisar antara -1,00 sampai dengan 1,00. Semakin tinggi indeks diskriminasi soal maka semakin mampu membedakan peserta didik yang pandai dengan yang kurang. Soal yang baik adalah yang memiliki kriteria daya pembeda yang minimal cukup. Rumus yang menentukan indeks diskriminasi (Daya Pembeda) adalah sebagai berikut:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Keterangan:

J_A = banyaknya peserta kelompok atas

J_B = banyaknya peserta kelompok bawah

B_A = banyaknya peserta kelompok atas yang menjawab benar

B_B = banyaknya peserta kelompok bawah yang menjawab benar

P_A = proporsi peserta kelompok atas yang menjawab benar

P_B = proporsi peserta kelompok bawah yang menjawab benar

Interpretasi Daya Pembeda dapat diklasifikasikan sebagai berikut:

D: 0,00-0,20 = jelek

D: 0,20-0,40 = cukup

D: 0,40-0,70 = baik

D: 0,70-1,00 = baik sekali

D: Negatif, semuanya tidak baik. Jadi semua butir soal yang mempunyai nilai D negatif sebaiknya dibuang saja.

(Arikunto, 2013)

Efektifitas Distraktor

Efektivitas pengecoh dapat diketahui melalui pola sebaran jawaban. Pola sebaran jawaban adalah distribusi *testee* dalam menentukan pilihan jawaban pada bentuk pilihan ganda. Cara memperolehnya adalah dengan menghitung banyaknya peserta tes yang memilih pilihan jawaban a,b,c,d,e atau yang tidak memilih apapun. Efektivitas pengecoh menunjukkan keberfungsian pengecoh secara baik atau tidak. Rumus untuk menghitung efektivitas pengecoh sebagai berikut:

$$IP = \frac{P}{\frac{(N - B)}{(n - 1)}} \times 100\%$$

Keterangan:

IP = indeks pengecoh.

P = jumlah peserta didik yang memilih pengecoh.

N = jumlah peserta didik yang ikut tes.

B = jumlah peserta didik yang menjawab benar.

n = jumlah alternatif jawaban.

1 = bilangan tetap

(Zainal, 2012)

Tafsiran terhadap hasil perhitungan setiap pengecoh pada satu butir soal dapat menggunakan kriteria sebagai berikut:

Sangat Baik	IP= 76%-125%
Baik	IP= 51%-75% atau 126%- 150%
Kurang Baik	IP= 26%-50% atau 151%-200%
Jelek	IP= 0%-25% atau 176%-200%
Sangat Jelek	IP= lebih dari 200%

Kemudian, hasil efektivitas pengecoh pada setiap butir soal diinterpretasikan menggunakan kriteria yang diadaptasi dari skala likert (Sugiyono, 2010)

- 1) Efektivitas pengecoh pada butir soal sangat baik apabila empat distraktor berfungsi.
- 2) Efektivitas pengecoh pada butir soal baik apabila tiga distraktor berfungsi.
- 3) Efektivitas pengecoh pada butir soal cukup apabila dua distraktor berfungsi.
- 4) Efektivitas pengecoh pada butir soal kurang baik apabila satu distraktor berfungsi.
- 5) Efektivitas pengecoh pada butir soal tidak baik apabila tidak ada distraktor yang berfungsi.

HASIL DAN PEMBAHASAN

Dalam pengembangan tes sumatif, yang perlu disiapkan pertama kali adalah kisi-kisi. Kisi-kisi dikembangkan mengikuti langkah-langkah dalam pengembangan tes sumatif yang telah ditentukan. Kemudian menyusun soal berdasarkan spesifikasi yang ada di dalam kisi-kisi tersebut. Ketika proses pengembangan soal perlu memperhatikan kaidah-kaidah dalam penulisan soal. Setelah mengikuti setiap langkah pengembangan soal hingga pada tahap penulisan tes pilihan ganda, maka akan dihasilkan 50 butir soal yang sesuai dengan aspek kognitif dengan sub aspek C1-C4 dan aspek psikomotor dengan sub aspek P1. Kemudian ditelaah kembali untuk meminimalisir kesalahan penulisan yang mungkin terjadi. Setelah itu soal diuji

cobakan ke kelas XI Bahasa SMA Negeri 1 Taman yang berjumlah sebanyak 30 peserta didik. Lembar Jawaban tes sumatif tersebut kemudian dianalisis secara kuantitatif.

Validitas

Validitas adalah suatu standart yang menunjukkan ketepatan suatu instrumen sehingga dapat mengukur apa yang seharusnya diukur. Berdasarkan analisis validitas dengan rumus korelasi point biserial menggunakan *Microsoft Excel 2010* diperoleh sebanyak 31 butir soal mempunyai r_{hitung} lebih dari r_{tabel} dengan $n=30$ ($df=28$) adalah 0,3610 dengan taraf signifikansi 5%. Sedangkan sisanya sebanyak 19 butir soal dinyatakan tidak valid.

Butir soal yang tidak valid memiliki beberapa penyebab yang mungkin terjadi yaitu butir soal yang kurang baik sehingga membuat peserta didik sukar memahami soal yang diujikan. Dalam hal ini perlu diadakannya perbaikan butir soal untuk mendapatkan soal yang valid. Selain itu, penyebab lainnya adalah jawaban peserta didik yang kurang baik sehingga pendidik perlu mengevaluasi sistem pembelajaran yang telah dilakukan agar peserta didik mampu menjawab soal dengan baik.

Reliabilitas

Berdasarkan hasil pengolahan data menggunakan *Microsoft Excel 2010*, nilai reliabilitas tes sumatif Bahasa Jerman kelas X Semester Ganjil sejumlah 50 butir adalah 0,88. Angka tersebut menunjukkan bahwa tes sumatif yang disusun memiliki tingkat keajekan atau konsistensi yang tinggi karena memiliki koefisien lebih besar dari 0,70.

Tingkat Kesukaran

Tingkat kesukaran dapat membedakan butir soal yang sukar atau tidak. Tingkat kesukaran dinyatakan dengan kategori mudah, sedang dan sukar. Kategori sukar memiliki indeks kesukaran (IK) antara 0,00--0,30, kategori sedang antara 0,30-

0,70 sedangkan kategori mudah antara 0,70-1,00. Pada hasil analisis, didapat 40 butir soal memiliki tingkat kesukaran yang mudah, 9 butir soal sedang dan 1 butir soal sukar. Dalam hal ini, butir soal yang baik adalah butir soal yang masuk dalam kategori sedang. Butir soal berkategori sedang menunjukkan bahwa butir soal tersebut tidak terlalu mudah dan tidak terlalu sukar.

Daya Pembeda

Daya pembeda setiap butir soal mampu menunjukkan perbedaan antara peserta didik yang telah menguasai materi dan yang belum. Daya pembeda dinyatakan dengan angka antara -1,00 sampai dengan 1,00. Semakin tinggi indeks daya pembeda maka semakin mampu membedakan peserta didik yang pandai maupun yang kurang. Interpretasi daya pembeda dapat diklasifikasikan dalam kategori jelek antara 0,00-0,20, kategori cukup antara 0,20-0,40, kategori baik antara 0,40-0,70 dan kategori baik sekali antara 0,70-1,00. Sedangkan daya pembeda yang menunjukkan angka negatif menunjukkan bahwa butir soal tidak baik dan sebaiknya dibuang saja.

Berdasarkan hasil analisis yang dilakukan pada 15 peserta didik kelompok atas (memiliki nilai tes tinggi) dan 15 peserta didik kelompok bawah (memiliki nilai tes rendah) didapatkan sebanyak 31 butir soal masuk dalam kategori jelek, sebanyak 10 butir soal masuk dalam kategori cukup dan sebanyak 9 butir soal masuk dalam kategori baik. Hasil tersebut menunjukkan bahwa banyak butir soal yang masuk dalam kategori jelek. Dalam hal ini butir soal yang berkategori jelek tidak bisa membedakan kemampuan peserta didik. Sehingga butir soal tersebut perlu evaluasi lebih lanjut.

Efektivitas Distraktor

Dari hasil analisis, terdapat 47 butir soal memiliki efektivitas distraktor yang tidak baik dan 3 butir soal memiliki efektivitas distraktor yang kurang baik. Hasil tersebut menunjukkan bahwa distraktor tersebut

tidak berfungsi, terlalu mencolok dan heterogen. Sehingga distraktor tersebut tidak memiliki daya tarik untuk dipilih oleh peserta didik yang kurang memahami materi. Butir soal tersebut sebaiknya ditindak lanjut dengan melakukan perbaikan atau revisi kemudian disimpan dalam bank soal untuk digunakan kembali pada pembuatan tes sumatif yang akan datang. Berikut cara membuat distraktor yang baik sehingga dapat mengecoh peserta didik (Surapranata, 2005), yaitu:

- 1) Pilihan jawaban dapat dimengerti peserta didik
- 2) Kata-kata yang dipilih merupakan kata-kata yang kedengarannya sama
- 3) Gunakan pilihan jawaban yang kemungkinan ada kaitannya
- 4) Gunakan bahasa baku agar distraktor tidak dilakukan lagi kebenarannya.

Berdasarkan hasil analisis kuantitatif di atas, butir soal yang dapat diterima berdasarkan kriteria sebanyak 20 butir soal. Sedangkan yang ditolak karena tidak memenuhi kriteria sebanyak 30 butir soal. Butir soal yang ditolak dapat dibuang atau direvisi berdasarkan jenis kriteria yang tidak terpenuhi. Berikut ini adalah rincian penyebab butir soal ditolak:

Tabel 4.7. Rincian Kriteria Butir Soal Ditolak

No.	Kriteria		No. Butir	Jumlah	Presentase
1	Validitas (Tidak Valid)		2, 5, 7, 9, 10, 13, 15, 22, 25, 34, 35, 36, 39, 42, 43, 45, 46, 49, 50	19	38 %
2	Tingkat Kesukaran	Mudah	1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 27, 28, 29, 30, 34, 35, 36, 37,	41	82 %

			38, 39, 40, 41, 42, 43, 45, 46, 48, 49, 50		
		S u k a r	8		
3	Daya Pembeda (jelek)		2, 5, 7, 8, 9, 10, 13, 15, 18, 19, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50	19	38 %
4	Efektivitas Distraktor	K u r a n g b a i k	16, 17, 48	50	100%
		t i d a k b a i k	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50		

Berdasarkan tabel diatas dapat diketahui bahwa faktor yang mempengaruhi ditolaknya butir soal adalah tingkat kesukaran dan efektivitas distraktor. Hal

tersebut menunjukkan bahwa tes sumatif masih terlalu mudah dan tidak memiliki distraktor yang berfungsi dengan baik. Selain itu masih banyak butir soal yang tidak valid dan belum mampu membedakan peserta didik yang memiliki kemampuan tinggi dan kemampuan rendah.

Ditinjau dari penyebab buruknya kualitas tes sumatif yang disusun, penyebab paling dominan adalah efektivitas pengecoh. Hal ini berkaitan dengan tingkat kesukaran yang cenderung mudah. Sehingga peserta didik yang menjawab tes mengetahui bahwa pengecoh bukan merupakan jawaban yang benar. Dalam hal ini salah satu faktor penyebab peserta didik mampu menjawab setia butir soal dengan mudah adalah jenjang peserta tes yang berada pada kelas XI. Sehingga hasil tingkat kesukaran tes sumatif akan berubah jika digunakan pada kelas X. Kemudian koefisien efektivitas pengecoh akan berubah dan diikuti dengan daya Pembedanya.

PENUTUP

SIMPULAN

Penelitian pengembangan tes sumatif bahasa Jerman kelas X Semester Ganjil di SMA Negeri 1 Taman dilakukan melalui beberapa tahapan untuk menjawab rumusan masalah dan mencapai tujuan penelitian dapat disimpulkan sebagai berikut:

1. Kisi-kisi tes sumatif pilihan ganda berdasarkan kurikulum 2013 dan taksonomi bloom revisi memungkinkan penyusunan butir soal yang bervariasi dalam aspek kognitif dan psikomotor. Kompetensi Dasar (KD) dan materi yang ditetapkan berdasarkan Kurikulum 2013 (K13). Indikator yang ditentukan berdasarkan aspek kognitif dengan sub aspek C1-C4 dan aspek psikomotor dengan sub aspek P1. Dengan kuantitas sebanyak 50 butir soal memiliki perbandingan tingkat kesukaran dengan

proporsi 3:5:2 berdasarkan *judgement* penyusun soal.

2. Berdasarkan analisis kuantitatif, sejumlah 31 butir tes valid atau sebanyak 72%, dengan tingkat reliabilitas sebesar 0,88 yang berarti tes sumatif yang disusun memiliki reliabilitas atau keajekan yang tinggi. Selain itu, sebesar 18% atau sebanyak 9 butir tes yang dikembangkan memiliki tingkat kesukaran sedang. Daya pembeda yang dimiliki tiap butir tes dalam kriteria cukup dan baik mencapai 62%. Namun tidak ada butir soal yang memiliki efektivitas distraktor yang cukup, baik dan sangat baik. Sehingga pada penelitian ini dihasilkan sebanyak 20 butir tes yang dapat diterima karena memenuhi kriteria yakni butir nomor; 1, 3, 4, 6, 11, 12, 14, 16, 17, 19, 21, 23, 24, 31, 32, 33, 40, 44, 47, 48 dan sebanyak 30 butir tes ditolak karena tidak memenuhi kriteria yakni butir nomor; 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50.

SARAN

Tes sumatif yang dikembangkan memiliki 20 butir soal yang diterima dan 30 butir soal yang ditolak. Ujian Akhir Semester Bahasa Jerman di SMA Negeri 1 Taman umumnya menggunakan Tes Sumatif berjumlah 50 butir selama 90 menit atau 2 jam pelajaran. Sehingga perlu adanya penelitian lanjut untuk memenuhi kriteria-kriteria butir soal yang dapat diterima berdasarkan kisi-kisi yang telah ditetapkan.

DAFTAR PUSTAKA

Abidin, Yunus. 2016. *Revitalisasi Penilaian Pembelajaran dalam Konteks Pendidikan Multiliterasi Abad ke-21*. Bandung: Refika Aditama.

Arikunto, Suharsimi. 2013. *Dasar-dasar Evaluasi Pendidikan (Edisi 2)*. Jakarta: Bumi Aksara.

Dudung, Agus. 2018. *Penilaian Psikomotor*. Depok: Karima

Giusto, Flavio Di, et al. 2018. *Multiple-Choice-Aufgaben (Teaching Guide for Higher & Professional Education)*. Winterthur: ZHAW School of Management and Law.

_____. 2014. *Kamus Besar Bahasa Indonesia*. Jakarta: Pusat Bahasa, Departemen Nasional Indonesia.

Kusaeri. 2014. *Acuan & Teknik Penilaian Proses & Hasil Belajar Dalam Kurikulum 2013*. Yogyakarta: Ar-Ruzz Media.

Kuswana, Wowo Sunaryo. 2012. *Taksonomi Kognitif*. Bandung: PT Remaja Rosdakarya Offset.

Nurghiyanoro, Burhan. 2010. *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: Universitas Negeri Yogyakarta (UNY)

Nuriyah, Nunung. 2014. *Evaluasi Pembelajaran: Sebuah Kajian Teori*. Jurnal Edueksos Vol III No. 1, Januari-Juni 2014. Tersedia di: <http://www.syekhnrjati.ac.id/jurnal/>

- index.php/edueksos/article/view/327/289. Diakses pada 03 Agustus 2019
- Nurwati, Andi. 2014. *Penilaian Ranah Psikomotor Siswa dalam Pelajaran Bahasa*. *Edukasia: Jurnal Penelitian Pendidikan Islam* Vol. 9, No. 2. Tersedia di: <http://journal.stainkudus.ac.id/index.php/Edukasia/article/viewFile/781/749>. Diakses pada 08 Agustus 2019.
- Purnomo, Edy. 2016. *Dasar-Dasar dan Perancangan Evaluasi Pembelajaran*. Yogyakarta: Media Akademi.
- Rosidin, Undang. 2017. *Evaluasi dan Asesmen Pembelajaran*. Yogyakarta: Media Akademi.
- Setyosari, Punaji. 2012. *Metode Penelitian Pendidikan dan Pengembangan*, Jakarta: Prenadamedia
- Subiyati, M. 1986, Validitas Cloze Test Sebagai Alat Ukur Kemampuan Membaca Bahasa Inggris, Laporan Penelitian, P2 IKIP Yogyakarta.
- Sugiyono. 2012. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Sukiman. 2017. *Sistem Penilaian Pembelajaran*. Yogyakarta: Media Akademi.
- Sunarti & Rahmawati, Selly. 2014. *Penilaian Kurikulum 2013 Membantu Guru dan Calon Guru Mengetahui Langkah-Langkah Penilaian Pembelajaran*. Yogyakarta: ANDI
- Trianto. 2007. *Model Pembelajaran Inovaif Berorientasi Konstruktivistik*. Surabaya: Pustaka Ilmu
- Zainal Arifin. 2012. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.

ENTWICKLUNG EINES SUMMATIVEN SPRACHENTESTS IM UNGERADES SEMESTER FÜR KLASSE X SMA NEGERI 1 TAMAN

Mas Uliatul Hikmah

Pädagogik der Deutsche Sprache, Fakultät für Sprache und Kunst, Universität Negeri
Surabaya

mashikmah16020094034@mhs.unesa.ac.id

Drs. H. Benny Herawanto Soesetyo, M.Psi.

Pädagogik der Deutsche Sprache, Fakultät für Sprache und Kunst, Universität Negeri
Surabaya

AUSZUG

Evaluierung ist ein Prozess des Sammelns von Fakten, um einen Überblick über den Lernfortschritt der Schüler zu geben. Um die Bewertungsziele zu erreichen, ist ein geeignetes Instrument erforderlich, da schlechte Bewertungsinstrumente dazu führen, dass die Bewertungsergebnisse nicht der Realität entsprechen. Bewertung des Deutschlernens in SMA Negeri 1 Taman, von denen einer eine summative test ist, die am Ende eines jeden Semesters durchgeführt wird. Bei der Entwicklung eines guten und rechenschaftspflichtigen Tests müssen Validität, Reliabilität, Schwierigkeitsgrad, Unterscheidungskraft und Wirksamkeit des Distraktors analysiert werden (Nurdiyanto, 2010). Damit ist die Formulierung des Problems diese Untersuchung, wie die Form der summativen Tests der deutschen Sprachklasse X SMA Negeri 1 Taman den Kriterien entspricht. Das Ziel diese Untersuchung ist es, einen summativen Test der Deutschklasse X SMA Negeri 1 Taman zu erstellen, der den Kriterien entspricht. Die vorgegebenen Kriterien sind a) die entwickelten Gegenstände sind valid, b) die entwickelten Gegenstände sind reliabel, c) der Schwierigkeitsgrad ist mäßig, c) die Unterscheidungskraft der Gegenstände ist minimal ausreichend und e) der Distraktor ist wirksam. Nach diesen Kriterien können Elemente akzeptiert werden, wenn sie mindestens drei Kriterien erfüllen, einschließlich validität und reliabilität. Die Artikel werden abgelehnt, wenn sie nicht alle Kriterien erfüllen. Um ein Bewertungsinstrument zu entwickeln, ist es notwendig, eine Reihe von Schritten in der Entwicklung eines vorbestimmten summativen Tests zu durchlaufen. Damit können summative Testbewertungsinstrumente erhalten werden, die der Qualität zugerechnet werden können. Diese Untersuchung gehört zu quantitative Untersuchung mit Versuchuntersuchung. Die Bevölkerung in diese Untersuchung ist die Schüler von Klasse XI SMA Negeri 1 Taman, die die deutsche Sprache als interessenübergreifendes und obligatorisches Fach aufnahm. Während die Stichprobe in diese Untersuchung ist die Schüler von Klasse XI SMA Negeri 1 Taman ist. Die Ergebnisse diese Untersuchung sind ein summatives Multiple-Choice-Testraster, das auf der überarbeiteten Taxonomie der kognitiven und psychomotorischen Aspekte von 2013 basiert. Von den 50 entwickelten Elementen kann die quantitative Analyse des Tests bis zu 20 Elemente akzeptieren, da sie die Kriterien der Nummern 1, 3, 4, 6, 11, 12, 14, 16, 17 erfüllen. 19, 21, 23, 24, 31, 32, 33, 40, 44, 47 und 48. Und bis zu 30 Artikel wurden abgelehnt, weil sie die Kriterien der Nummern 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50 nicht erfüllten.

Schlüsselwort : Testentwicklung, Deutscher Summative Test, Ungerades Semester

ABSTRACT

Evaluation is a process of collecting facts to give an overview of students' learning progress. To achieve the goals of the evaluation, an appropriate instrument is needed because poor evaluation instruments will result in the assessment results being not by reality. Evaluation of German learning in SMA Negeri 1 Taman one of which is a summative test that is carried out at the end of each semester. In developing a good and accountable test, it is necessary to analyze the validity, reliability, level of difficulty, differentiation and effectiveness of the distractor (Nurgiyantoro, 2010). So that the formulation of the problem in this study is how the form of summative tests of German language class X SMA Negeri 1 Taman that fits the criteria. The purpose of this study is to produce a summative test of German language class X SMA Negeri 1 Taman that fits the criteria. The predetermined criteria are a) the items developed are valid, b) the items developed are reliable, c) the level of difficulty is moderate c) the distinguishing power of items is minimally sufficient, and e) the distractor is effective. From these criteria, items can be accepted if they meet a minimum of three criteria including validity and reliability. The items are rejected if they do not meet all the criteria. To develop an evaluation instrument it is necessary to go through a series of steps in the development of a predetermined summative test. So that it can be obtained summative test evaluation instruments that can be accounted for quality. This research uses quantitative research with the type of trial research. The population in this study was class XI SMA Negeri 1 Taman, which took German language as a cross-interest and compulsory subject. While the sample in this study was class XI SMA Negeri 1 Taman. The results of this study are a multiple-choice summative test grid based on the 2013 curriculum and bloom revised taxonomy of cognitive and psychomotor aspects. Of the 50 items that have been developed, the quantitative analysis of the test shows as many as 20 test items can be accepted because it meets the criteria of numbers 1, 3, 4, 6, 11, 12, 14, 16, 17, 19, 21,23, 24, 31, 32, 33, 40, 44, 47, and 48. And as many as 30 items were rejected because they did not meet the criteria of number 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50.

Keywords : Test Development, German Summative Test, Odd Semester

Universitas Negeri Surabaya

HINTERGRUND

Bei der Evaluierung werden Daten gesammelt, um einen Überblick über den Lernfortschritt der Schüler zu erhalten. Beschreibungen dieser Entwicklungen müssen den Pädagogen bekannt sein, um festzustellen, ob die Schüler den Lernprozess richtig erlebt haben oder nicht. Dann können Pädagogen geeignete Alternativen bei der Entscheidungsfindung bestimmen. Um die Bewertungsziele zu erreichen, ist ein geeignetes Instrument erforderlich, da schlechte Bewertungsinstrumente dazu führen, dass die Bewertungsergebnisse nicht der Realität entsprechen. Deutsch in SMA Negeri 1 Taman ist ein Fach, das von runden mit Schwerpunkt Sprache und Pflichtfach belegt werden muss. Um die Beschreibung des Lernfortschritts der Schüler herauszufinden, führen die Pädagogen Auswertungen in Form von summativen Tests durch, die am Ende jedes Semesters mit Multiple-Choice-Tests durchgeführt werden. Nach Nurgiyantoro (2010, 26) wurden die Tests bei der Entwicklung guter und verantwortlicher Tests auf ihre Validität und Reliabilität, Schwierigkeitsgrad, Unterscheidungskraft und die Wirksamkeit des Distraktors hin analysiert. Aus den Ergebnissen der Analyse werden empirische Informationen erhalten, die die Qualität des Tests belegen. Basierend auf den Anforderungen des obigen summativen Tests möchte der Erzieher einen summativen Test, der angemessen ist oder den Kriterien entspricht. Diese Untersuchung ist also eine Forschungsentwicklung für summative Tests, die speziell für die Klasse X (ungerades Semester) als Anfänger im Erlernen von Deutsch entwickelt wurden.

Die Formulierung des Problems in diese Untersuchung ist, wie die Form der summativen Tests der deutschen Sprachklasse X SMA Negeri 1 Taman die Kriterien erfüllt. Während der Zweck diese Untersuchung ist es, einen summativen

Test der deutschen X-Klasse SMA Negeri 1 Taman zu erstellen, der die Kriterien erfüllt. Artikel Artikelkriterien umfassen; a) die entwickelten Gegenstände sind valid, b) die entwickelten Gegenstände sind reliabel, c) der Schwierigkeitsgrad ist mäßig, c) die Unterscheidungskraft der Gegenstände ist minimal ausreichend und e) der Distraktor ist wirksam. Basierend auf den oben genannten Kriterien werden Elemente wie folgt akzeptiert oder abgelehnt: 1) Artikel werden akzeptiert, wenn sie mindestens drei Kriterien einschließlich Validität und Reliabilität erfüllen, und 2) Artikel werden abgelehnt, wenn sie nicht alle Kriterien erfüllen.

Um Missverständnisse in diese Untersuchung zu vermeiden, sind die folgenden Einschränkungen zu beachten.

1. Das Instrument wird in Form von summativen Testfragen in Form von Multiple-Choice-Tests präsentiert, die auf der Grundlage von Lehrplan und KD in K13 entwickelt werden
2. Testfragen werden durch Zuweisen von Fragenrastern entwickelt.
3. Bei der in diese Untersuchung genannten Entwicklung handelt es sich um die Herstellung summativer Tests, die auf den Schritten der Entwicklung summativer Tests basieren.
4. Datenverarbeitung mit dem Microsoft Excel 2010-Datenverarbeitungsprogramm.
5. Summative Tests werden an Schülern der Klasse XI in SMA Negeri 1 Taman durchgeführt, da sie das gesamte Material der Klasse X im ungeraden Semester erhalten haben.

METHODE

Diese Forschung verwendet einen quantitativen Forschungsansatz. Ein quantitativer Ansatz wird verwendet, um empirische Daten zur Qualität von summativen Tests aus der Arbeit der Studenten zu erhalten. Empirische Daten

in Form von Zahlen werden dann mithilfe des Microsoft Excel-Programms mithilfe eines Computers analysiert. Die Art der Forschung, die in diese Untersuchung verwendet wird, ist eine Versuchsforschung. Die Bevölkerung in diese Untersuchung ist Schüler der Klasse XI von SMA Negeri 1 Taman, die Deutsch als interessenübergreifendes und obligatorisches Fach belegten. Während die Stichprobe in diese Untersuchung die Klasse XI Sprache ist, unter Berücksichtigung, dass Deutsch ein Pflichtfach für die Sprachabteilung ist.

Die Datenanalysetechniken in diese Untersuchung verwendeten eine quantitative Analyse durch Messen mehrerer Variablen, nämlich der Validität des Gegenstands, der Reliabilität des Tests, des Schwierigkeitsindex, des Unterscheidungskraftindex und der Wirksamkeit des Distraktors. Bei der Analyse dieser quantitativen Daten mit der Microsoft Excel-Anwendung nach der Formel der einzelnen Variablen.

Validität

Die Validität misst die Genauigkeit dessen, was gemessen werden soll. Die Artikel Validität misst die Genauigkeit des Artikels mit dem gemessenen Aspekt. Der Weg, um Punkte für die Validität eines Artikels zu erhalten, besteht darin, den Korrelationsindex zwischen den Bewertungen jedes Artikels und den Gesamtbewertungen zu berechnen. Ein Punkt gilt als valid, wenn die Punktzahl der Gesamtpunktzahl entspricht. Es besteht also eine positive Korrelation zwischen Item-Scores und Gesamt-Scores. Die Validität von Multiple-Choice-Elementen wurde unter Verwendung der Biserialpunktkorrelation berechnet. Anschließend wird eine Konsultation mit rtable mit einem Synergiegrad von 5% gemäß der Anzahl der untersuchten Studenten durchgeführt. Der Artikel gilt als valid, wenn $rpbi > rtable$, wenn $rpbi$ kleiner oder gleich $rtable$ ist, dann ist der Artikel invalid. Die Validität von

Multiple-Choice-Formularelementen kann mithilfe der Biserialpunkt-Korrelationsformel berechnet werden:

$$r_{pbi} = \frac{M_p - M_t}{S_t} \sqrt{\frac{p}{q}}$$

Information:

$rpbi$ = Biserieller Korrelationskoeffizient

M_p = durchschnittliche Punktzahl des Themas, das für Elemente, deren Validität gesucht wird, richtig antwortet

M_t = durchschnittliche Gesamtpunktzahl

S_t = Standardabweichung der Gesamtpunktzahl des Unternehmens

p = Anteil der Schüler, die richtig geantwortet haben

p = (Anzahl der Schüler, die richtig geantwortet haben)

$(p = \frac{\text{banyaknya siswa yang menjawab benar}}{\text{jumlah seluruh peserta didik}})$

q = Anteil der Schüler, die falsch geantwortet haben ($q = 1-p$)

(Arikunto, 2013)

Reliabilität

Die Reliabilität wird ausgeführt, um das Niveau der Testergebnisse oder das Niveau der Konsistenz zwischen Elementen zu bestimmen. Die Reliabilität des Problems kann durch Betrachten der Größe des Reliabilitätskoeffizienten des Problems erkannt werden. Die Reliabilität von Multiple-Choice-Fragen kann mithilfe der KR-20-Formel bestimmt werden. Das Problem gilt als Reliabilit, wenn der Fragensicherheitskoeffizient (r_{11}) gleich oder größer als 0,70 ist. Je höher der Koeffizient, desto Reliabilität ist das Problem. Die Reliabilität kann mit der KR-20-Formel wie folgt berechnet werden:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(\frac{S^2 - \Sigma pq}{S^2} \right)$$

Information:

r₁₁ = allgemeine Test Reliabilität

p = Anteil der Probanden, die den Gegenstand richtig beantwortet haben

q = der Anteil der Probanden, die den Gegenstand falsch beantwortet haben (q = 1-p)

Σpq = wie oft das Produkt von p und q ist

n = Stückzahl

S = Standardabweichung des Tests

(Arikunto, 2013)

Schwierigkeitsgrad

Der Schwierigkeitsgrad ist eine Gelegenheit, ein Problem mit einem bestimmten, in einem Index ausgedrückten Fähigkeitsgrad zu beantworten. Der Schwierigkeitsgrad wird in Form von Anteilen zwischen 0,00 und 1,00 ausgedrückt. Je kleiner der Schwierigkeitsindex ist, desto schwieriger ist das Problem. Gegenstandskugeln können als gute Gegenstände bezeichnet werden, wenn sie nicht zu schwierig und nicht zu einfach sind. Ein guter Gegenstand hat also einen mittleren oder mittleren Schwierigkeitsgrad mit einem Schwierigkeitsgrad zwischen 0,30 und 0,70. Schwierigkeitsindexnummern können mit der folgenden Formel erhalten werden:

$$P = \frac{B}{JS}$$

Information:

P = Artikelschwierigkeitsindexnummer

B = Anzahl der Testpersonen, die den fraglichen Gegenstand richtig beantworten können

JS = Anzahl der Testpersonen, die den Lernleistungstest ablegen

Die Schwierigkeitsindexkriterien lauten wie folgt:

0,00-0,30 = schwierige Kategoriefragen

0,30-0,70 = Fragen mittlerer Kategorie

0,70-1,00 = einfache Kategoriefragen

(Arikunto, 2013)

Unterscheidungskraft

Die Unterscheidungskraft kann den Unterschied zwischen Studenten aufzeigen, die das Material mit weniger beherrschen. Die Fähigkeit zur Unterscheidung von Fragen kann anhand der Größe des Index der Fragendiskriminierung ermittelt werden. Der Artikelunterscheidungsindex reicht von -1,00 bis 1,00. Je höher der Index für die Diskriminierung von Fragen ist, desto besser können intelligente Schüler von weniger unterschieden werden. Eine gute Frage ist eine, die ausreichend minimale Unterscheidungskriterien aufweist. Die Formel, die den Unterscheidungsindex (Unterscheidungsvermögen) bestimmt, lautet wie folgt:

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} = P_A - P_B$$

Information:

JA = Anzahl der Teilnehmer in der oberen Gruppe

JB = Anzahl der Teilnehmer in der unteren Gruppe

BA = Anzahl der Top-Gruppenteilnehmer, die richtig geantwortet haben

BB = die Anzahl der Teilnehmer in der unteren Gruppe, die richtig geantwortet haben

PA = Anteil der Top-Gruppenteilnehmer, die richtig geantwortet haben

PB = Anteil der Teilnehmer der unteren Klasse, die richtig geantwortet haben

Die unterscheidenden Interpretationen können wie folgt klassifiziert werden:

D: 0,00-0,20 = schlecht

D: 0,20 - 0,40 = genug

D: 0,40-0,70 = gut

D: 0,70-1,00 = sehr gut

D: Negativ, alles ist nicht gut. Daher sollten alle Elemente mit einem negativen D-Wert verworfen werden.

(Arikunto, 2013)

Wirksamkeit von Distraktoren

Die Täuschungswirksamkeit kann durch das Muster der Antwortverteilung erkannt werden. Das Muster der Antwortverteilung ist die Verteilung des Probanden bei der Bestimmung der Wahl der Antworten in der Multiple-Choice-Form. Der Weg dorthin führt über die Anzahl der Testteilnehmer, die die Antwortmöglichkeiten a, b, c, d, e gewählt haben oder die nichts gewählt haben. Die Wirksamkeit der Täuschung zeigt an, ob die Täuschung richtig funktioniert oder nicht. Die Formel zur Berechnung der Täuschungswirksamkeit lautet wie folgt:

$$IP = \frac{P}{\frac{(N - B)}{(n - 1)}} \times 100\%$$

Information:

IP = Täuschungsindex.

P = Anzahl der Schüler, die sich für eine Täuschung entschieden haben.

N = Anzahl der Schüler, die an der Prüfung teilnehmen.

B = Anzahl der Schüler, die richtig geantwortet haben.

n = Anzahl der alternativen Antworten.

1 = feste Nummer

(Zainal, 2012)

Die Interpretation der Ergebnisse der Berechnung jedes Trickers für einen Artikel kann nach folgenden Kriterien erfolgen:

Sehr gute IP = 76% -125%

Entweder IP = 51% -75% oder
126% - 150%

Schlechte IP = 26% -50% oder

151% -200%

Schlechte IP = 0% -25% oder

176% -200%

Sehr schlechte IP = mehr als 200%

Dann wurden die Ergebnisse der Wirksamkeit von Betrug bei jedem Gegenstand unter Verwendung von Kriterien interpretiert, die von der Likert-Skala angepasst wurden (Sugiyono, 2010).

- 1) Die Wirksamkeit der Täuschung im Gegenstand ist sehr gut, wenn vier Ablenker funktionieren.
- 2) Die Wirksamkeit der Täuschung auf dem Gegenstand ist gut, wenn drei Ablenker funktionieren.
- 3) Die Wirksamkeit der Täuschung im Gegenstand ist ausreichend, wenn zwei Ablenker funktionieren.
- 4) Die Wirksamkeit der Täuschung im Gegenstand ist nicht gut, wenn ein Ablenker funktioniert.
- 5) Die Wirksamkeit der Täuschung in der Sache ist nicht gut, wenn es keinen funktionierenden Auftragnehmer gibt.

ERGEBNISSE UND DISKUSSION

Bei der Entwicklung von summativen Tests, was zuerst im Raster vorbereitet werden muss. Das Gitter wird nach den Schritten bei der Entwicklung eines vorbestimmten summativen Tests entwickelt. Ordnen Sie dann die Fragen anhand der Angaben im Raster an. Bei der Entwicklung von Fragen müssen die Regeln beim Verfassen von Fragen beachtet werden. Nach jedem Schritt der Fragenentwicklung bis zum Schreiben von Multiple-Choice-Tests werden 50 Fragen generiert, die kognitiven Aspekten mit Unteraspekten von C1-C4 und psychomotorischen Aspekten mit Unteraspekten von P1 entsprechen. Überprüfen Sie die Informationen dann

erneut, um mögliche Schreibfehler zu minimieren. Danach wurden die Fragen in der Klasse XI Sprache von SMA Negeri 1 Taman mit insgesamt 30 Schülern getestet. Das summative Testantwortblatt wird dann quantitativ analysiert.

Validität

Die Validität ist ein Standard, der die Genauigkeit eines Instruments angibt, damit gemessen werden kann, was gemessen werden soll. Basierend auf der Validitätsanalyse mit der Biseriapunkt-Korrelationsformel unter Verwendung von Microsoft Excel 2010 haben 31 Elemente mehr als rtable mit $n = 30$ ($df = 28$) gezählt, was einem Signifikanzniveau von 5% entspricht. Während die restlichen 19 Artikel für invalid erklärt wurden.

Invalid Artikel Artikel haben mehrere mögliche Ursachen, nämlich, dass der Artikel nicht gut ist und es den Schülern daher schwer fällt, die zu testende Frage zu verstehen. In diesem Fall muss das Element verbessert werden, um Valid Fragen zu erhalten. Außerdem sind die Antworten der Schüler andere Ursachen, die nicht gut sind. Daher müssen die Pädagogen das von den Schülern durchgeführte Lernsystem bewerten, um die Fragen gut beantworten zu können.

Reliabilität

Basierend auf den Ergebnissen der Datenverarbeitung mit Microsoft Excel 2010 beträgt der Reliabilität des summativen Tests der Deutschsprachklasse X Odd Semester von 50 Items 0,88. Dieser Wert gibt an, dass der kompilierte summative Test ein hohes Maß an Konsistenz oder Konsistenz aufweist, da sein Koeffizient größer als 0,70 ist.

Schwierigkeitsgrad

Der Schwierigkeitsgrad kann schwierige Gegenstände unterscheiden oder nicht. Der Schwierigkeitsgrad wird in einfachen, mittleren und schwierigen Kategorien

angegeben. Die harte Kategorie hat einen Schwierigkeitsindex (IK) zwischen 0,00 und 0,30, die mittlere Kategorie zwischen 0,30 und 0,70, während die leichte Kategorie zwischen 0,70 und 1,00 liegt. In den Ergebnissen der Analyse wurde festgestellt, dass 40 Gegenstände einen leichten Schwierigkeitsgrad hatten, 9 Gegenstände mäßig waren und 1 Gegenstand schwierig war. In diesem Fall sind gute Artikel Artikel, die in die mittlere Kategorie fallen. Die Artikel in der Kategorie zeigen, dass der Artikel nicht zu einfach und nicht zu schwierig ist.

Unterscheidungskraft

Die Unterscheidungskraft jedes Gegenstands kann den Unterschied zwischen Schülern, die das Material beherrschen, und solchen, die es nicht beherrschen, aufzeigen. Das Unterscheidungsvermögen wird als Zahl zwischen -1,00 und 1,00 ausgedrückt. Je höher der Leistungsindex ist, desto besser können kluge und weniger kluge Schüler unterschieden werden. Die Interpretation der Unterscheidungskraft kann in die schlechte Kategorie zwischen 0,00-0,20, die ausreichende Kategorie zwischen 0,20-0,40, die gute Kategorie zwischen 0,40-0,70 und die ausgezeichnete Kategorie zwischen 0,70-1,00 eingeteilt werden. Während die Unterscheidungskraft, die eine negative Zahl anzeigt, anzeigt, dass der Gegenstand nicht gut ist und weggeworfen werden sollte.

Basierend auf den Ergebnissen der Analyse, die an 15 Schülern in der oberen Gruppe (mit einem hohen Testergebnis) und 15 Schülern in der unteren Gruppe (mit einem niedrigen Testergebnis) durchgeführt wurde, wurden 31 Punkte in die schlechte Kategorie eingeschlossen, 10 Punkte eingeschlossen in der ausreichenden Kategorie und 9 Artikel in beiden Kategorien enthalten. Diese Ergebnisse weisen darauf hin, dass viele Elemente in der Kategorie "Schlecht" enthalten sind. In diesem Fall können die

als schlecht eingestuften Elemente die Fähigkeiten der Schüler nicht unterscheiden. Die Items müssen also noch weiter bewertet werden.

Wirksamkeit von Distraktoren

Aus den Ergebnissen der Analyse geht hervor, dass 47 Artikel eine schlechte Ablenkungswirksamkeit aufweisen und 3 Artikel eine schlechte Ablenkungswirksamkeit aufweisen. Diese Ergebnisse weisen darauf hin, dass der Distraktor nicht funktioniert, zu auffällig und heterogen ist. Der Distraktor hat also nicht die Anziehungskraft, die von Schülern ausgewählt werden muss, die das Material nicht verstehen. Nach diesen Elementen sollten Verbesserungen oder Überarbeitungen vorgenommen und dann in der Fragendatenbank gespeichert werden, um sie für zukünftige summative Tests erneut zu verwenden. So machen Sie einen guten Distraktor, damit er die Schüler überlisten kann (Surapranata, 2005):

1) Die Wahl der Antworten kann von den Schülern verstanden werden

2) Die gewählten Wörter sind Wörter, die gleich klingen

3) Verwenden Sie die Antwortmöglichkeiten, die im Zusammenhang stehen könnten

4) Verwenden Sie die Standardsprache, damit der Distraktor nicht mehr von der Wahrheit ausgeführt wird.

Basierend auf den Ergebnissen der obigen quantitativen Analyse können 20 Elemente anhand von Kriterien akzeptiert werden. Diejenigen, die abgelehnt wurden, weil sie die Kriterien nicht erfüllten, betrugen 30 Punkte. Abgelehnte Elemente können aufgrund der Art der nicht erfüllten Kriterien verworfen oder überarbeitet werden. Im Folgenden finden Sie Einzelheiten dazu, warum die Artikel abgelehnt wurden:

Tabelle 4.7. Detaillierte Kriterien für abgelehnte Artikel

Nr.	Kriterien		Nr. Artikelfrage	Anzahl	Prozent
1	Validität (Invalid)		2, 5, 7, 9, 10, 13, 15, 22, 25, 34, 35, 36, 39, 42, 43, 45, 46, 49, 50	19	38 %
2	Schwierigkeitsgrad	e i i f a c h	1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 25, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 48, 49, 50	41	82 %
		s c h w i e r i g	8		
3	Unterscheidungskraft (schlecht)		2, 5, 7, 8, 9, 10, 13, 15, 18, 19, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50	19	38 %
4	die Wirksamkeit des	Weniger gut	16, 17, 48	50	100%

	Distrak- tors	n i c h t g u t	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 49, 50		
--	------------------	--	---	--	--

Aus der obigen Tabelle geht hervor, dass die Faktoren, die die Zurückweisung des Artikels beeinflussen, der Schwierigkeitsgrad und die Wirksamkeit des Distraktors sind. Dies zeigt, dass summative Tests immer noch zu einfach sind und keine funktionierenden Distraktoren haben. Außerdem gibt es immer noch viele Gegenstände, die invalid sind und nicht in der Lage waren, Schüler mit hohen und niedrigen Fähigkeiten zu unterscheiden.

Gemessen an den Ursachen für die schlechte Qualität der zusammengestellten summativen Tests ist die Wirksamkeit der Betrüger die Hauptursache. Dies hängt mit dem Schwierigkeitsgrad zusammen, der in der Regel einfach ist. Studenten, die den Test beantworten, wissen, dass Täuschung nicht die richtige Antwort ist. In diesem Fall ist einer der Faktoren, die die Schüler dazu veranlassen, die Fragen leicht zu beantworten, die Anzahl der Testteilnehmer in Klasse XI. Die Ergebnisse des Schwierigkeitsgrads der summativen Prüfung ändern sich also, wenn sie in Klasse X verwendet werden. Dann ändert sich der Wirkungsgrad der Täuschung und es folgt die Unterscheidungskraft.

SCHLUSS

ZUSAMMENFASSUNG

Die Erforschung der Entwicklung summativer Sprachtests für die Klasse X Odd Semester in SMA Negeri 1 Taman erfolgt in mehreren Schritten, um die Problemstellung zu beantworten und die Forschungsziele zu erreichen.

1. Der summative Multiple-Choice-Test auf der Grundlage des Lehrplans 2013 und der überarbeiteten Blütentaxonomie ermöglicht die Zusammenstellung von Elementen, die sich in kognitiven und psychomotorischen Aspekten unterscheiden. Basiskompetenz (KD) und Material basierend auf dem Lehrplan 2013 (K13). Die Indikatoren werden basierend auf kognitiven Aspekten mit den Unteraspekten C1-C4 und psychomotorischen Aspekten mit den Unteraspekten P1 bestimmt. Bei einer Stückzahl von 50 ergibt sich nach Einschätzung des Fragestellers ein Schwierigkeitsgrad von 3: 5: 2.

2. Basierend auf quantitativer Analyse sind viele 31 Testobjekte valid oder sogar zu 62%, mit einem Reliabilitätsgrad von 0,88, was bedeutet, dass die vorbereiteten summativen Tests eine hohe Reliabilität aufweisen. Außerdem hatten 18% oder 9 der entwickelten Tests einen mäßigen Schwierigkeitsgrad. Die Unterscheidungskraft, die jeder Gegenstand in den Kriterien besitzt, ist ausreichend und gut, um 72% zu erreichen. Es gibt jedoch keine Artikel mit ausreichender, guter und sehr guter Ablenkungswirkung. Damit wurden in diese Untersuchung bis zu 20 Prüflinge hergestellt, die akzeptiert werden können, weil sie die Kriterien erfüllen, nämlich Artikelnummer; 1, 3, 4, 6, 11, 12, 14, 16, 17, 19, 21, 23, 24, 31, 32, 33, 40, 44, 47, 48 und bis zu 30 Testobjekte wurden abgelehnt, weil sie erfüllte nicht die Kriterien, dh Artikelnummer; 2, 5, 7, 8, 9, 10, 13, 15, 18, 20, 22, 25, 26, 27, 28, 29, 30, 34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46, 49, 50.

VORSCHLAGEN

Der entwickelte summative Test hat 20 eingegangene und 30 zurückgewiesene Punkte. Die Deutsch-Abschlussprüfung bei SMA Negeri 1 Taman besteht in der Regel aus einem 50-Punkte-Summentest für 90 Minuten oder 2 Stunden. Daher sind weitere Untersuchungen erforderlich, um die Kriterien für Elemente zu erfüllen, die auf der Grundlage des angegebenen Gitters akzeptiert werden können.

BIBLIOGRAPHIE

Abidin, Yunus. 2016. *Revitalisasi Penilaian Pembelajaran dalam Konteks Pendidikan Multiliterasi Abad ke-21*. Bandung: Refika Aditama.

Arikunto, Suharsimi. 2013. *Dasar-dasar Evaluasi Pendidikan (Edisi 2)*. Jakarta: Bumi Aksara.

Dudung, Agus. 2018. *Penilaian Psikomotor*. Depok: Karima

Giusto, Flavio Di, et al. 2018. *Multiple-Choice-Aufgaben (Teaching Guide for Higher & Professional Education)*. Winterthur: ZHAW School of Management and Law.

_____. 2014. *Kamus Besar Bahasa Indonesia*. Jakarta: Pusat Bahasa, Departemen Nasional Indonesia.

Kusaeri. 2014. *Acuan & Teknik Penilaian Proses & Hasil Belajar Dalam Kurikulum 2013*. Yogyakarta: Ar-Ruzz Media.

Kuswana, Wowo Sunaryo. 2012. *Taksonomi Kognitif*. Bandung: PT Remaja Rosdakarya Offset.

Nurgiyantoro, Burhan. 2010. *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: Universitas Negeri Yogyakarta (UNY)

Nuriyah, Nunung. 2014. *Evaluasi Pembelajaran: Sebuah Kajian Teori*. Jurnal Edueksos Vol III No. 1, Januari-Juni 2014. Tersedia di: <http://www.syekhnurjati.ac.id/jurnal/index.php/edueksos/article/view/327/289>. Diakses pada 03 Agustus 2019

Nurwati, Andi. 2014. *Penilaian Ranah Psikomotor Siswa dalam Pelajaran Bahasa*. *Edukasia: Jurnal Penelitian Pendidikan Islam* Vol. 9, No. 2. Tersedia di: <http://journal.stainkudus.ac.id/index.php/Edukasia/article/viewFile/781/749>. Diakses pada 08 Agustus 2019.

Purnomo, Edy. 2016. *Dasar-Dasar dan Perancangan Evaluasi Pembelajaran*. Yogyakarta: Media Akademi.

Rosidin, Undang. 2017. *Evaluasi dan Asesmen Pembelajaran*. Yogyakarta: Media Akademi.

Setyosari, Punaji. 2012. *Metode Penelitian Pendidikan dan Pengembangan*, Jakarta: Prenadamedia

Subiyati, M. 1986, Validitas Cloze Test Sebagai Alat Ukur Kemampuan Membaca Bahasa Inggris, Laporan Penelitian, P2 IKIP Yogyakarta.

Sugiyono. 2012. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.

Sukiman. 2017. *Sistem Penilaian Pembelajaran*. Yogyakarta: Media Akademi.

Sunarti & Rahmawati, Selly. 2014. *Penilaian Kurikulum 2013 Membantu Guru dan Calon Guru Mengetahui Langkah-Langkah Penilaian Pembelajaran*. Yogyakarta: ANDI

Trianto. 2007. *Model Pembelajaran Inovatif Berorientasi Konstruktivistik*. Surabaya: Pustaka Ilmu

Zainal Arifin. 2012. *Evaluasi Pembelajaran*. Bandung: PT Remaja Rosdakarya.