

BAB III

METODOLOGI PENELITIAN

A. Metode penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian dan pengembangan (*Research and Development*). Metode ini digunakan karena penelitian ini menghasilkan produk berupa tes diagnostik membaca BIPA dan menguji keefektifan tes diagnostik tersebut. Hal ini sejalan dengan pendapat Sugiono (2011, hlm 297) bahwa metode penelitian dan pengembangan digunakan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut.

Penelitian ini dilakukan dengan mengacu pada asumsi dasar berikut.

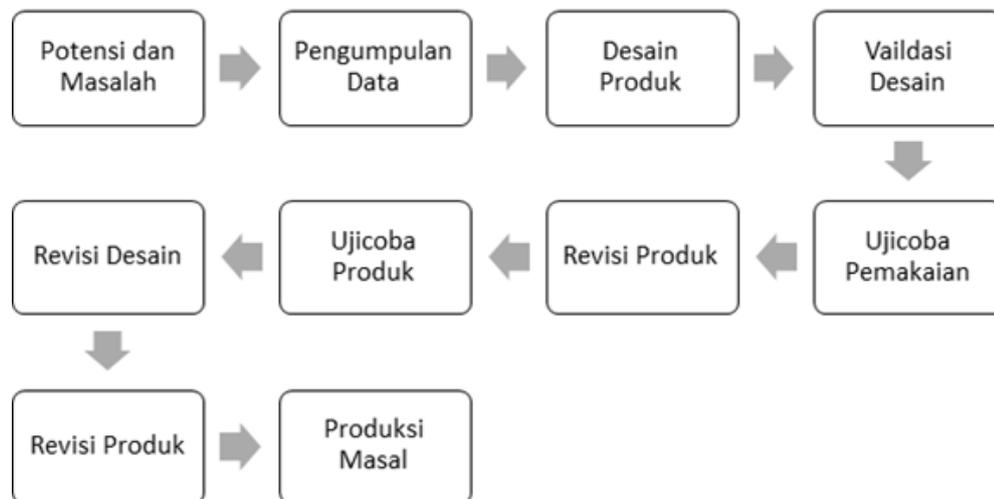
1. Tes diagnostik membaca berbasis *dynamic assessment* sebagai alternatif alat evaluasi BIPA dapat menilai keterampilan membaca pembelajar BIPA dengan tepat.
2. Butir soal dikategorikan sebagai soal yang baik setelah melalui analisis tingkat kesulitan, daya beda, dan analisis pengecoh.
3. Alat evaluasi membaca yang tepat memiliki validitas, realibilitas, dan kepraktisan yang signifikan.

B. Desain penelitian

Desain penelitian dalam *Research and Development* bermacam-macam, seperti desain *Borg and Gall*, desain *Dick and Carey*, desain *Hannafin and Peck*. Ketiga desain penelitian tersebut memiliki persamaan dalam tahapan-tahapan awal. Perbedaan ketiga desain tersebut terlihat pada tahapan akhir dalam pengembangan dan pengimplementasian produk yang dihasilkan dalam penelitian.

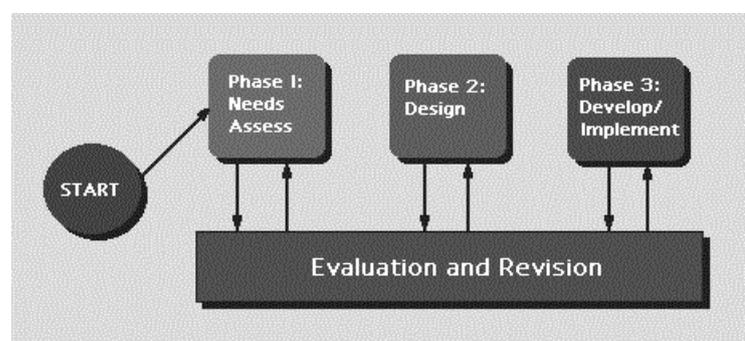
Sugiono (2011) memaparkan penelitian yang mengadaptasi desain *Borg and Gall* dilakukan dengan sepuluh langkah. Langkah pertama dimulai dengan mengidentifikasi potensi dan masalah penelitian kemudian melakukan pengumpulan informasi sebagai bahan untuk merancang produk untuk mengatasi masalah yang ditemukan sebelumnya. Setelah itu, desain produk dilakukan untuk

membuat produk awal. Desain tersebut divalidasi oleh para ahli kemudian direvisi jika ada yang perlu diperbaiki. Produk awal yang dikembangkan berdasarkan desain tersebut diujicoba dengan dua tahap, yaitu tahap ujicoba produk dan ujicoba pemakaian. Revisi pun kembali dilakukan apabila dalam kedua tahap tersebut ditemukan hal yang belum sesuai dan perlu perbaikan. Setelah semua proses tersebut dilakukan, produksi masal pun dilakukan.



Gambar 3.1 Desain penelitian *Borg and Gall* Sumber: Sugiono (2011)

Model Hannafin dan Peck adalah desain pengembangan yang terdiri dari pada tiga fase yaitu fase analisis keperluan, fase desain, fase pengembangan dan implementasi (Hannafin dan Peck, dalam Affandi dan Badarudin, 2011 dan Majid Abdul, 2011). Dalam model ini, penilaian dan pengulangan perlu dijalankan dalam setiap fase. Model ini lebih berorientasi produk melalui tiga fase berikut.



Gambar 3.2 Desain Hannafin dan Peck Sumber: www.futureu.net

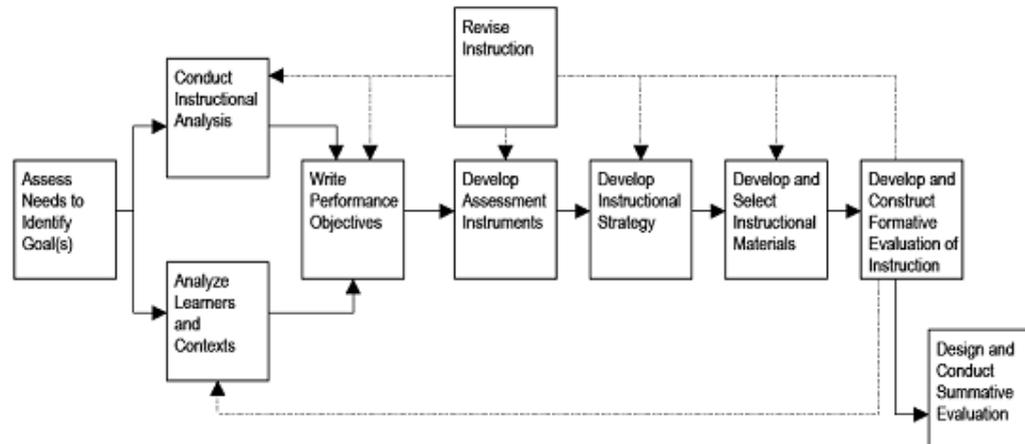
Fase pertama ini dilakukan dengan mengidentifikasi kebutuhan-kebutuhan dalam mengembangkan suatu produk pembelajaran. Fase pertama dalam penelitian ini dilakukan dengan menganalisis karakteristik alat evaluasi membaca yang dibutuhkan pembelajar BIPA. Setelah semua kebutuhan diidentifikasi, penilaian terhadap hasil itu perlu dilakukan sebelum melanjutkan ke tahap fase desain.

Hasil analisis kebutuhan dari fase pertama dipindahkan ke dalam bentuk desain tes diagnosis membaca BIPA yang akan menjadi landasan pengembangan produk tes diagnosis tersebut. Fase desain bertujuan untuk mengidentifikasi dan mendokumentasikan kaidah yang paling baik untuk mencapai tujuan pembuatan produk tersebut. Desain produk yang telah selesai dirancang perlu diberikan penilaian oleh para ahli. Penilaian tersebut akan menjadi bahan revisi untuk memperbaiki desain produk tersebut.

Desain alat evaluasi membaca BIPA dari fase kedua dikembangkan menjadi produk tes diagnosis membaca BIPA. Tahap pengembangan dan implementasi produk tersebut dilakukan dengan penilaian formatif dan penilaian sumatif. Hannafin dan Peck (1988) menyatakan penilaian formatif adalah penilaian yang dilakukan sepanjang proses pengembangan produk sedangkan penilaian sumatif dilakukan setelah produk tersebut selesai dikembangkan. Oleh karena itu, tahap pengembangan dilakukan dengan penilaian formatif oleh para ahli terhadap produk tes diagnosis membaca BIPA. Setelah itu, tahap implementasi dilakukan dengan penilaian sumatif, yaitu mengujicobakan produk yang telah direvisi berdasarkan penilaian formatif kepada pembelajar BIPA. Jika hasil uji coba menunjukkan adanya perbaikan, maka revisi pun akan dilakukan untuk membuat produk akhir.

Dick dkk (2009) mendeskripsikan desain *Dick and Carey* dalam sepuluh tahapan, *identify instructional goal(s)*, *conduct instructional analysis*, *analysis learners and contexts*, *write performance objectives*, *develop assessment instrument*, *develop instructional strategy*, *develop and select instructional materials*, *design and conduct formative evaluation of instruction*, *revise instruction*, dan *design and conduct summative evaluation*. Implementasi tahapan

awal penelitian dalam desain penelitian ini hampir sama dengan desain sebelumnya. Perbedaannya terletak pada evaluasi sumatif yang dilakukan dengan dua tahap, yaitu *one-to-one evaluation with learners* dan *small group evaluation*. Kedua tahapan tersebut dilakukan sebelum tahap *judgement expert*.



Gambar 3.3 Desain penelitian *Dick and Carey* Sumber: *ete.cltt.ubc.ca*

Desain pengembangan yang digunakan dalam penelitian ini mengadaptasi desain *Dick and carey*. Namun, tahap *one-to-one evaluation with learners* dan *small group evaluation* pada tahap penilaian formatif tidak dilakukan. Hal ini disebabkan keterbatasan partisipan yang bersedia mengikuti tahap implementasi sehingga tidak dapat melakukan uji coba pemakaian dan produksi masal.

C. Partisipan

Partisipan yang terlibat dalam penelitian ini adalah tiga puluh pembelajar BIPA dari berbagai negara yang dipilih secara acak.

Tabel 3.1 Daftar partisipan penelitian

No	Nama	Asal Negara
1	Adilah Salaeh	Thailand
2	Arnon	Thailand
3	Atif Bensulong	Thailand
4	Baek Hyun	Korea Selatan
5	Fatma Bahaa Eldin	Mesir
6	Hakimi	Pakistan

Hani Maryana, 2015

PENGEMBANGAN TES DIAGNOSTIK KEMAMPUAN MEMBACA BERBASIS ASESMEN DINAMIK SEBAGAI ALTERNATIF ALAT EVALUASI BIPA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

7	Husna	Malaysia
8	Jahirulislam	Thailand
9	Jang Seung So	Korea Selatan
10	Kyoung Sun Seo	Korea Selatan
11	Lee Dong Ki	Korea Selatan
12	Mee Jung Jo	Korea Selatan
13	Mizuki	Jepang
14	Muna	Thailand
15	Naemah	Thailand
16	Nagisa	Jepang
17	Nara	Jepang
18	Narim	Thailand
19	Nor Hannah	Malaysia
20	Rowiyah	Thailand
21	Samaritdinov Zukhriddin	Uzbekistan
22	Siti Humaira	Malaysia
23	Suhaimi	Thailand
24	Sulaiman	Thailand
25	Susrina Kalee	Thailand
26	Wanabdullah Paduka	Thailand
27	Wanerfam	Thailand
28	Wiam Huda	Thailand
29	Yuna Yajima	Jepang
30	Zhong Xiao Wen	China

D. Instrumen penelitian

Instrumen atau alat pengumpul data yang dipergunakan dalam penelitian ini, yaitu observasi, timbangan para ahli atau *judgement expert*, dan soal tes.

1. Pedoman wawancara

Wawancara yang dilakukan dengan semi terstruktur dengan praktisi BIPA dari berbagai lembaga penyelenggara BIPA, yaitu yaitu lembaga BIPA di universitas (UPI, UGM, UNJ, UI), tempat kursus (Wisma Bahasa Yogyakarta), Pusat Bahasa dan Balai Bahasa Bandung. Wawancara ini bertujuan untuk mengetahui berbagai jenis alat evaluasi dan instrumen tes yang dipergunakan di lembaga tersebut. Wawancara ini dilakukan pada penelitian pendahuluan.

2. Pedoman observasi

Observasi dilakukan pada tahap penelitian pendahuluan pada pemelajar BIPA. Observasi dilaksanakan selama kegiatan tutor di balai bahasa UPI selama tiga puluh jam. Pemelajar yang diobservasi berjumlah tiga orang dari negara yang berbeda. Mereka adalah Mia dari Pakistan, Sejin dari Korea Selatan, dan Hedi dari Afganistan. Mia dan Sejin berada pada tingkat dasar satu dan Hedi berada pada tingkat dasar dua. Observasi awal ini bertujuan untuk melihat karakteristik kemampuan membaca ketiga pemelajar BIPA tersebut.

3. Timbangan para ahli

Format timbangan pakar dilakukan untuk menilai draft awal instrumen tes diagnostik membaca. Para ahli menilai instrumen tes yang berjumlah sembilan puluh soal dengan rentang skor 1-5. Semakin besar skor yang diberikan, maka tingkat validitas soal tersebut semakin tinggi. Kolom komentar dan saran disediakan untuk merevisi butir-butir soal yang belum mencapai skor maksimal. Timbangan pakar ini bertujuan untuk menilai validitas soal sehingga layak untuk diujicobakan kepada pemelaar BIPA.

4. Tes

Instrumen penelitian yang terakhir berupa instrumen soal dalam tes diagnostik membaca BIPA. Tes diagnostik ini memuat sembilan puluh soal yang mengukur kemampuan membaca pemelajar BIPA. Tes ini terdiri dari lima wacana yang berbeda tingkat keterbacaannya, mulai dari sangat mudah, mudah, menengah, sulit, dan sangat sulit. Setiap wacana terdiri dari delapan belas soal yang mewakili kemampuan kognitif dalam *cognitive factors in reading process*' Grabe. Tes ini dilakukan secara online dengan bantuan perangkat lunak *google form* kepada tiga puluh pemelajar BIPA dari berbagai negara.

E. Prosedur Penelitian

Kronologis langkah-langkah penelitian yang dilakukan sesuai dengan desain penelitian yang telah dioperasionalkan sebelumnya adalah sebagai berikut.

1. Penelitian pendahuluan

Penelitian pendahuluan bertujuan untuk analisis kebutuhan tes diagnostik membaca BIPA. Hal ini dilakukan untuk membuktikan bahwa penelitian mengenai tes diagnostik membaca untuk pemelajar asing merupakan inovasi baru dalam dunia ke-BIPA-an sehingga menarik untuk dilakukan. Penelitian pendahuluan ini dilakukan untuk mengungkapkan alasan yang disertai fakta mengenai ketertaikan dalam penelitian tes diagnostik ini. Penelitian pendahuluan ini dilakukan untuk merealisasikan tahapan *identify instructional goal(s)* dalam desain *Borg and Gall*.

2. Rancangan tes diagnostik membaca BIPA

Tahapan ini dilakukan untuk merealisasikan *conduct instructional analysis, analysis learners and contexts, write performance objective*. Tahapan ini dilakukan dengan mengidentifikasi parameter tes diagnostik membaca BIPA. Parameter tes diagnostik menggunakan teori Grabe mengenai untuk pengembangan instrumen soal dan formula LIX untuk mengukur keterbacaan soal. Setelah itu, perancangan draft awal dilakukan berdasarkan parameter tersebut.

3. Pengembangan tes diagnostik membaca BIPA

Tahapan ini dilakukan dengan mengembangkan draft awal yang sudah dirancang menjadi produk awal berupa instrumen tes. Produk awal ini kemudian divalidasi oleh para ahli. Jika hasil validasi para ahli menunjukkan instrumen belum layak diimplementasikan, maka revisi produk awal tersebut dilakukan supaya produk tersebut menunjukkan validasi yang tinggi dan layak untuk diuji coba pada tahap implementasi. Tahapan ini dilakukan untuk merealisasikan langkah-langkah *develop assessment instrument, develop instructional strategy, develop and select instructional materials*.

4. Implementasi tes diagnostik membaca BIPA

Tahap ini dilakukan untuk merealisasikan *design and conduct formative evaluation of instruction, revise instruction*, dan *design and conduct summative evaluation* dalam desain penelitian *Borg and Gall*. Hal ini dilakukan dengan mengujicobakan instrumen tes kepada pemelajar BIPA. Setelah itu, penilaian formatif dan sumatif dilakukan untuk menghasilkan tes diagnostik membaca BIPA yang memiliki validitas dan realibilitas yang tinggi.

F. Teknik analisis data

Data dalam penelitian ini diolah dengan berbagai jenis teknik, yaitu *judgmental method of improving assesment, empirical method of improving assesment*, validitas, dan realibilitas. Analisis timbangan pakar dalam pengembangan instrumen tes diagnostik dilakukan dengan membandingkan hasil penilaian ketiga pakar sehingga instrumen tes diagnostik membaca BIPA layak untuk diuji coba kepada pemelajar BIPA.

1. *Judgmental Method of Improving Assesment*

Metode timbangan ini dilakukan pada fase pengembangan. Para ahli BIPA diminta untuk menilai produk tes diagnosis membaca dengan berpedoman pada lima kriteria Popham (2011) berikut.

- a. Kesesuaian penulisan soal dengan pedoman penulisan soal secara umum dan pedoman penulisan bentuk-bentuk soal.
- b. Ketepatan kontribusi skor penilaian.
- c. Keakuratan instrumen soal dengan materi pembelajaran.
- d. Kesesuaian instrumen soal dengan materi pembelajaran.
- e. Keadilan proporsi materi pembelajaran dalam instrumen soal.

2. *Empirical Method of Improving Assesment*

Metode empirik ini dilakukan pada fase implementasi. Metode ini diadaptasi dari Popham (2011) dengan mengolah hasil data empirik pembelajar BIPA dalam merespon tes diagnosis membaca. Data tersebut akan dianalisis berdasarkan tingkat kesulitan, daya pembeda, dan analisis pengecoh.

Metode empirik pengembangan instrumen dilakukan dengan mengolah hasil data empirik pemelajar BIPA dalam merespon instrumen tes diagnostik membaca BIPA. Metode empirik ini menilai tingkat kesulitan, daya pembeda, dan analisis pengecoh.

Analisis tingkat kesulitan bertujuan untuk mengetahui seberapa sulit instrumen soal yang diberikan. Analisis ini dilakukan dengan menghitung nilai- p dengan rumus berikut.

$$\text{Tingkat Kesulitan } P = \frac{R}{T}$$

Keterangan:

- R = Banyaknya siswa yang menjawab benar soal tersebut
- T = Jumlah siswa yang menjawab soal tersebut

Semakin tinggi nilai p , semakin banyak siswa yang menjawab benar soal tersebut. Semakin rendah nilai p , semakin sedikit siswa yang menjawab salah soal tersebut.

Analisis daya pembeda soal bertujuan untuk melihat kemampuan butir soal dalam membedakan siswa yang mampu dan tidak mampu dalam merespon instrumen soal dengan benar. Analisis ini dapat dilakukan dengan empat tahapan berikut.

- Susun lembar jawaban dari skor yang tertinggi sampai skor terendah
- Bagilah lembar jawaban tersebut menjadi kelompok tinggi (P_h) dan kelompok rendah (P_l). Banyaknya lembar jawaban harus sama di setiap kelompok.
- Hitung p value masing-masing kelompok.
- Hitung *Discrimination index* (D) setiap soal dengan rumus:

$$D = P_h - P_l$$

Setelah nilai *Discrimination index* (D) diperoleh, koversikan nilai tersebut dalam Pedoman Evaluasi Daya Pembeda Soal berikut.

Tabel 3.2 Interpretasi indeks diskriminasi

Indeks Diskriminasi	Evaluasi Soal
.40 - ke atas	Soal yang sangat baik
.30 - .39	Soal yang baik, tapi ada kemungkinan untuk diperbaiki
.20 - .29	Soal yang kurang baik, biasanya perlu perbaikan
.19 – ke bawah	Soal yang tidak baik, harus diganti atau direvisi

Analisis pengecoh dimaksudkan untuk menilai fungsi pengecoh masing-masing pilihan jawaban tersebut. Selain itu, tingkat kesulitan dan daya pembeda juga sebagai faktor penentu kualitas pilihan jawaban masing-masing soal tersebut.

Tabel 3.3 Analisis pengecoh

No Soal	Pilihan jawaban	R	P	D	Pengecoh
	Pilihan jawaban	>0	>0	>0	+
	Pilihan jawaban	>0	>0	≥0	-
	Pilihan jawaban	0	>0	∞	-
	Pilihan jawaban				*

Keterangan:

R : jumlah siswa yang memilih pilihan jawaban tersebut

P : tingkat kesulitan soal

D : daya pembeda

Pengecoh

tanda (+) : pengecoh berfungsi dengan baik

tanda (-) : pengecoh tidak berfungsi

tanda (*) : kunci jawaban soal

Kualitas pengecoh dalam pilihan jawaban ditentukan oleh banyaknya pemelajar yang memilih pilihan jawaban tersebut, tingkat kesulitan soal, dan daya pembeda soal. Pengecoh bertanda (+) menandakan pengecoh pilihan jawaban tersebut berfungsi dengan baik. Hal ini ditandai dengan adanya siswa yang memilih pilihan jawaban tersebut, dan tingkat kesulitan serta daya pembeda soal bernilai positif. Pengecoh bertanda (-) menandakan pengecoh pilihan jawaban tersebut tidak berfungsi. Hal ini ditandai dengan tidak adanya pemelajar yang memilih pilihan jawaban tersebut, atau daya pembeda soal bernilai negatif atau kurang atau sama dengan nol. Jika dalam kolom pengecoh, pilihan jawaban bertanda (*) maka pilihan jawaban tersebut merupakan kunci jawaban soal.

Validitas soal dilakukan pada tahap pengembangan produk. Validitas ini dilakukan dengan penilaian para ahli dari tiga aspek validitas, yaitu isi, kriteria, dan konstruk. Hal ini sesuai dengan pernyataan Djiwandono (2011) berikut.

Validitas dapat dikenali dan dibuktikan keberadaannya melalui kajian terhadap kesesuaian isi (validitas isi), kesesuaian kriteria (validitas kriteria), dan kesesuaian konstruk dalam asrti konsep, kerangka teori, atau dasar pemikiran (validitas konstruk).

Realibilitas instrumen tes diagnostik dibuktikan dengan perhitungan statistika dengan perhitungan manual dan bantuan perangkat lunak minitab. Realibilitas instrumen tes diagnostik dihitung dengan realibilitas belah dua. Tahapan pertama dalam perhitungan belah dua yaitu membagi tes menjadi dua bagian, yaitu tes A

Hani Maryana, 2015

**PENGEMBANGAN TES DIAGNOSTIK KEMAMPUAN MEMBACA BERBASIS ASESMEN DINAMIK
SEBAGAI ALTERNATIF ALAT EVALUASI BIPA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

dan tes B. Tes A merupakan tes yang memuat soal-soal genap, sedangkan tes B memuat soal-soal ganjil. Setelah itu, skor ketiga puluh siswa dihitung kembali berdasarkan pembagian tes tersebut. Tahap selanjutnya menghitung koefisien korelasi antara dua bagian tes dengan rumus Pearson product moment berikut.

Setelah mendapatkan koefisien korelasi kedua belah tes tersebut, realibilitas tes seutuhnya dapat dihitung dengan rumus berikut.

$$r_{xx} = \frac{n \times r}{(n - 1)r - 1}$$

Keterangan:

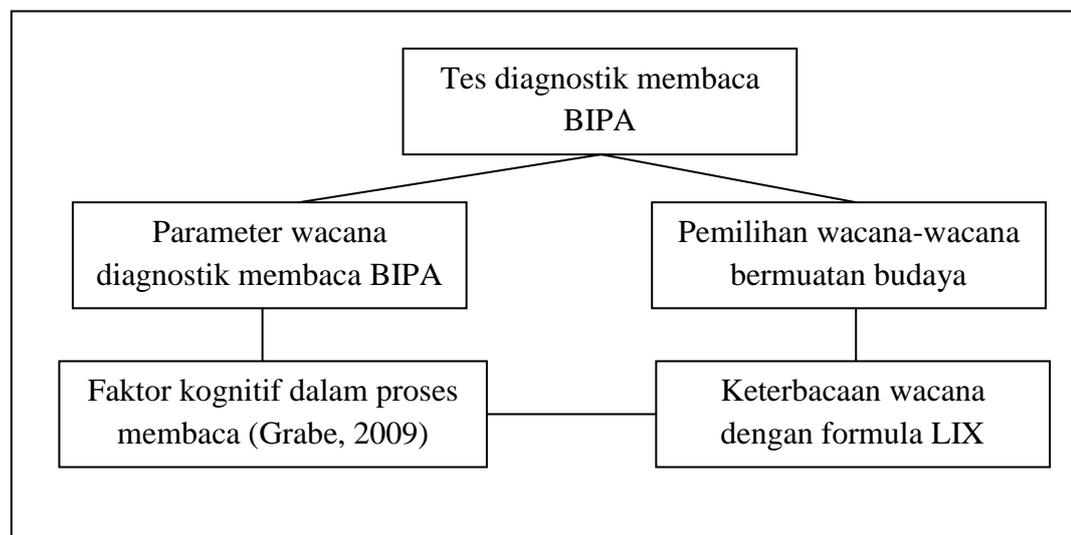
r_{xx} : realibilitas tes seutuhnya

r : koefisien korelasi antara ke dua bagian tes, hasil perhitungan *Pearson product moment*

n : jumlah belahan tes

G. Skema produk awal

Skema rancangan produk awal tes diagnostik membaca BIPA ini adalah sebagai berikut.



Gambar 3.4 Skema produk awal

Rancangan produk awal tes diagnostik membaca BIPA ini dikembangkan dengan menggunakan parameter Grabe mengenai faktor kognitif dalam membaca proses. Proses membaca tersebut terbagi menjadi dua, yaitu proses membaca

tingkat rendah dan tingkat tinggi. Setiap proses membaca tersebut dijabarkan menjadi tiga indikator.

Wacana-wacana yang dipilih dalam produk awal ini merupakan wacana-wacana yang mengandung nilai-nilai budaya Indonesia sehingga pemelajar BIPA. Hal ini bertujuan agar pemelajar BIPA dapat memahami budaya masyarakat di Indonesia. Wacana yang dipilih berasal dari berbagai sumber kemudian diukur keterbacaannya dengan menggunakan formula LIX untuk mendapatkan lima wacana yang dapat merepresentasikan semua tingkat kesulitan wacana, mulai wacana yang sangat mudah sampai wacana yang sangat sulit.

Soal-soal yang dikembangkan mengacu pada pemahaman kelima wacana tersebut. Setiap wacana terdiri atas tiga puluh soal. Soal-soal tersebut mewakili setiap indikator. Indikator-indikator tersebut dikembangkan menjadi tiga soal. Hal ini dilakukan untuk menyasati keterbatasan waktu partisipan dalam tahap ujicoba. Selain itu, hal ini juga dilakukan untuk memilih satu soal yang paling tepat untuk merepresentasikan setiap indikator tersebut sehingga dapat menghasilkan produk akhir instrumen tes yang memiliki nilai validitas, realibilitas, dan kepraktisan yang tinggi.