

ABSTRAK

Penelitian ini bertujuan untuk menghasilkan instrumen penilaian berbasis HOTS (*High Order Thinking Skill*) berbentuk uraian yang layak berdasarkan kriteria kualitas instrumen penilaian, diantaranya yaitu validitas, reliabilitas, taraf kesukaran, dan daya pembeda serta memenuhi validasi dari segi konten, konstruktivitas, dan bahasa. Penelitian ini menggunakan jenis penelitian pengembangan dengan model penelitian *formative research* yang terdiri dari tahap *preliminary*, *self evaluation*, *prototyping*, dan *field test*. Instrumen dalam penelitian ini berupa angket validasi *expert review*, angket respon siswa, dan tes. Hasil analisis data dari tahap *prototyping* bagian *expert review* dan *one-to-one* memperoleh soal HOTS dengan kategori “sangat valid”, namun dilakukan perubahan berdasarkan saran dan komentar dari keduanya dan menghasilkan *prototype* 2. Pada tahap uji coba *one-to-one* dilakukan revisi kembali pada penggunaan bahasa sehingga menghasilkan *prototype* 3. Kemudian berdasarkan hasil dari tahap *field test*, *prototype* 3 diujikan dan dianalisis validitas menghasilkan 3 soal tidak valid dan 7 soal valid dan memiliki nilai reliabilitas 0.83 dengan interpretasi sangat tinggi. Pada analisis taraf kesukaran menghasilkan 6 soal sedang dan 4 soal mudah, sedangkan pada analisis daya pembeda menghasilkan 5 soal baik dan 5 soal kurang baik.

Kata kunci: instrumen penilaian, *high order thinking skill*, *formative research*

ABSTRACT

This study aimed to produce HOTS-based assessment instruments (High Order Thinking Skill) in the form of appropriate descriptions based on the quality criteria of assessment instruments, including validity, reliability, degree of difficulty, and differentiation as well as fulfilling validation in terms of content, constructs, and languages. This study used a type of development research with a formative research model consisting of the preliminary stage, self-evaluation, prototyping, and field tests. The instruments in this study were questionnaire validation expert reviews, student response questionnaires, and tests. Results from the prototyping stage, the expert review, and one-to-one section received HOTS questions with the category "very valid", but made changes based on suggestions and comments from both of them and produced a prototype 2. At the one-to-one trial phase revisions were made to the use of language so that it produces a prototype 3. Then based on the results of the field test stage, prototype 3 was tested and analyzed validity produces 3 invalid questions and 7 valid questions and has a reliability value of 0.83 with very high interpretation. On the analysis, the difficulty level produces 6 intermediate questions and 4 easy questions, while the distinguishing analysis produces 5 good questions and 5 poor questions.

Keywords: *assessment instruments, high order thinking skills, formative research*