In [2]:

```
#pip install numpy
#pip install pandas
#pip install matplotlib
#pip install seaborn
```

In [6]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [7]:

```
df = pd.read_csv("student_Score.csv")
```

In [8]:

```
print(df.head())
```

```
   Unnamed: 0  Gender EthnicGroup          ParentEduc     LunchType TestPrep  \
0           0  female         NaN   bachelor's degree      standard     none   
1           1  female     group C        some college      standard      NaN   
2           2  female     group B     master's degree      standard     none   
3           3    male     group A  associate's degree  free/reduced     none   
4           4    male     group C        some college      standard     none   

  ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings TransportMeans  \
0             married     regularly          yes         3.0     school_bus   
1             married     sometimes          yes         0.0            NaN   
2              single     sometimes          yes         4.0     school_bus   
3             married         never           no         1.0            NaN   
4             married     sometimes          yes         0.0     school_bus   

  WklyStudyHours  MathScore  ReadingScore  WritingScore  
0            < 5         71            71            74  
1         05-Oct         69            90            88  
2            < 5         87            93            91  
3         05-Oct         45            56            42  
4         05-Oct         76            78            75  
```

In [11]:

```
#description of all data min, max, count, etc
df.describe()
```

Out[11]:

	Unnamed: 0	NrSiblings	MathScore	ReadingScore	WritingScore
count	30641.000000	29069.000000	30641.000000	30641.000000	30641.000000
mean	499.556607	2.145894	66.558402	69.377533	68.418622
std	288.747894	1.458242	15.361616	14.758952	15.443525
min	0.000000	0.000000	0.000000	10.000000	4.000000
25%	249.000000	1.000000	56.000000	59.000000	58.000000
50%	500.000000	2.000000	67.000000	70.000000	69.000000
75%	750.000000	3.000000	78.000000	80.000000	79.000000
max	999.000000	7.000000	100.000000	100.000000	100.000000

In [12]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30641 entries, 0 to 30640
Data columns (total 15 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Unnamed: 0           30641 non-null  int64  
 1   Gender               30641 non-null  object 
 2   EthnicGroup          28801 non-null  object 
 3   ParentEduc           28796 non-null  object 
 4   LunchType            30641 non-null  object 
 5   TestPrep             28811 non-null  object 
 6   ParentMaritalStatus  29451 non-null  object 
 7   PracticeSport        30010 non-null  object 
 8   IsFirstChild         29737 non-null  object 
 9   NrSiblings           29069 non-null  float64
 10  TransportMeans       27507 non-null  object 
 11  WklyStudyHours       29686 non-null  object 
 12  MathScore            30641 non-null  int64  
 13  ReadingScore         30641 non-null  int64  
 14  WritingScore         30641 non-null  int64  
dtypes: float64(1), int64(4), object(10)
memory usage: 3.5+ MB
```

In [13]:

```python
#calculating how much values are null
df.isnull().sum()
```

Out[13]:

```
Unnamed: 0                0
Gender                    0
EthnicGroup            1840
ParentEduc             1845
LunchType                 0
TestPrep               1830
ParentMaritalStatus    1190
PracticeSport           631
IsFirstChild            904
NrSiblings             1572
TransportMeans         3134
WklyStudyHours          955
MathScore                 0
ReadingScore              0
WritingScore              0
dtype: int64
```

In [18]:

```python
#drop unnamed column
```

In [15]:

```python
df= df.drop("Unnamed: 0", axis = 1)
print(df.head())
```

```
  Gender EthnicGroup          ParentEduc     LunchType TestPrep  \
0  female         NaN   bachelor's degree      standard     none   
1  female     group C        some college      standard      NaN   
2  female     group B     master's degree      standard     none   
3    male     group A  associate's degree  free/reduced     none   
4    male     group C        some college      standard     none   

  ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings TransportMeans  \
0             married     regularly          yes         3.0     school_bus   
1             married     sometimes          yes         0.0            NaN   
2              single     sometimes          yes         4.0     school_bus   
3             married         never           no         1.0            NaN   
4             married     sometimes          yes         0.0     school_bus   

  WklyStudyHours  MathScore  ReadingScore  WritingScore  
0            < 5         71            71            74  
1         05-Oct         69            90            88  
2            < 5         87            93            91  
3         05-Oct         45            56            42  
4         05-Oct         76            78            75  
```

In [19]:
```
#transformation of table data
```

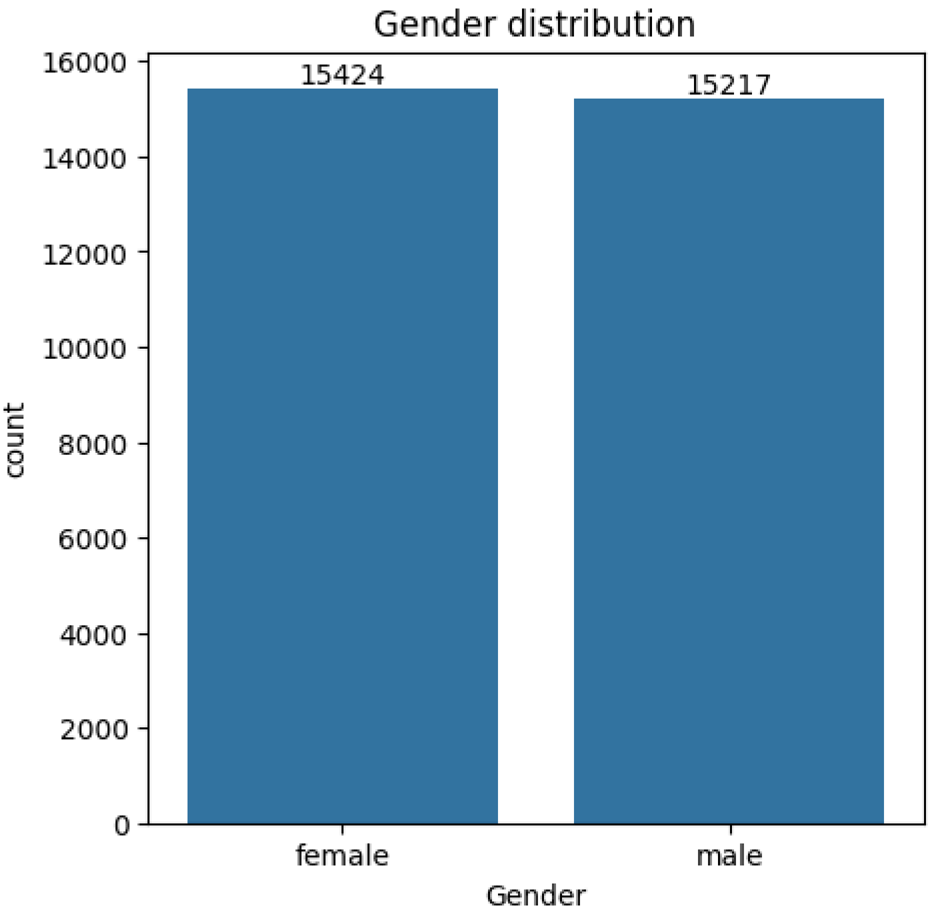
In [20]:
```
df["WklyStudyHours"] = df["WklyStudyHours"].str.replace("05-Oct","5-10")
df.head()
```

Out[20]:

	Gender	EthnicGroup	ParentEduc	LunchType	TestPrep	ParentMaritalStatus	Practi
0	female	NaN	bachelor's degree	standard	none	married	ı
1	female	group C	some college	standard	NaN	married	so
2	female	group B	master's degree	standard	none	single	so
3	male	group A	associate's degree	free/reduced	none	married	
4	male	group C	some college	standard	none	married	so

In [21]:
```
#gender distribution
```

In [41]:
```
plt.figure(figsize= (5,5))
ax = sns.countplot(data = df, x = "Gender")
ax.bar_label(ax.containers[0])
plt.title("Gender distribution")
plt.show()
```

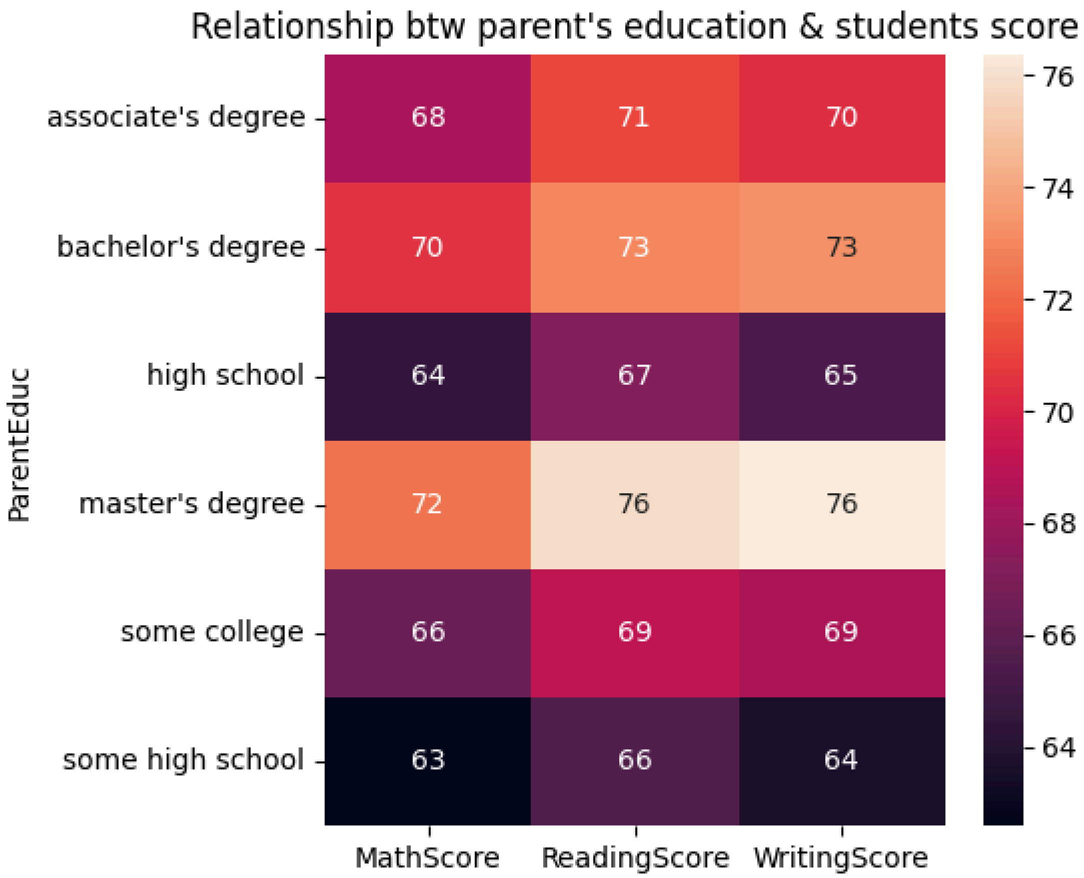


--from above charts, we have analysed that the no. of females are more then males

In [32]:
```
gb = df.groupby("ParentEduc").agg({"MathScore":'mean',  "ReadingScore":'mean',
print(gb)
```

```
                    MathScore  ReadingScore  WritingScore
ParentEduc
associate's degree  68.365586     71.124324     70.299099
bachelor's degree   70.466627     73.062020     73.331069
high school         64.435731     67.213997     65.421136
master's degree     72.336134     75.832921     76.356896
some college        66.390472     69.179708     68.501432
some high school    62.584013     65.510785     63.632409
```

In [42]:
```
plt.figure(figsize= (5,5))
sns.heatmap(gb, annot = True)
plt.title("Relationship btw parent's education & students score")
plt.show()
```

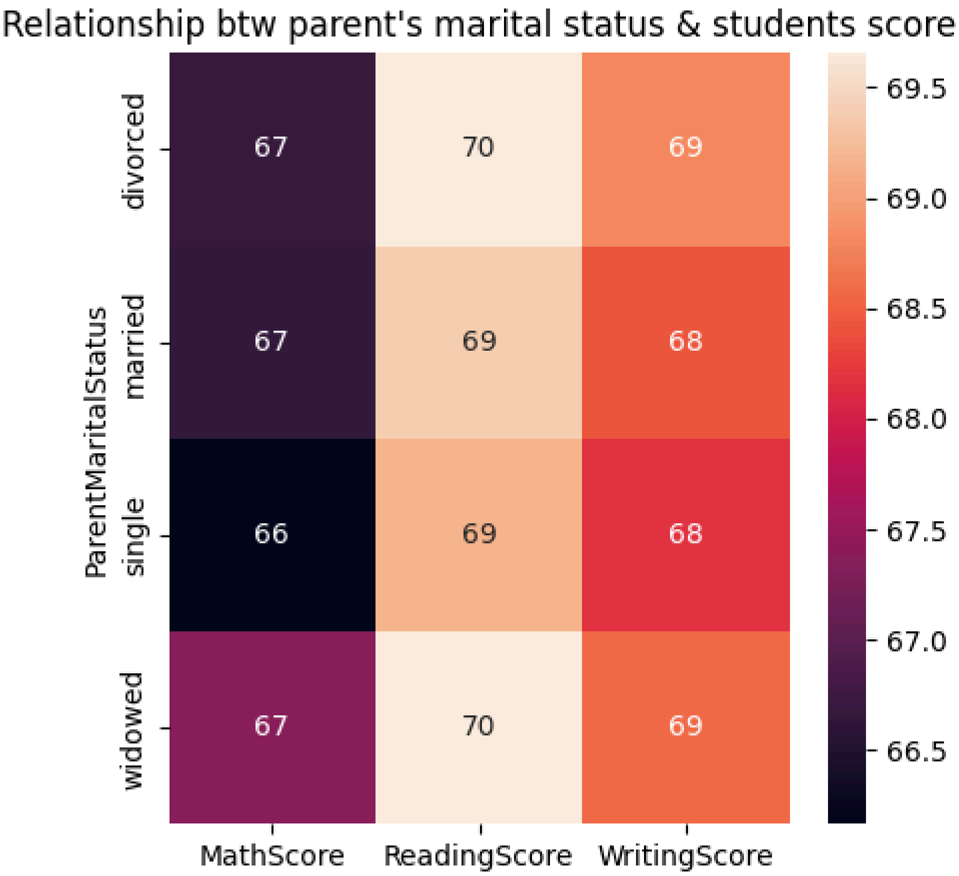


--From the above chart we have concluded that the education of parents have a good impact on there kids.

```
In [39]: gb1 = df.groupby("ParentMaritalStatus").agg({"MathScore":'mean',        "Reading
         print(gb1)
```

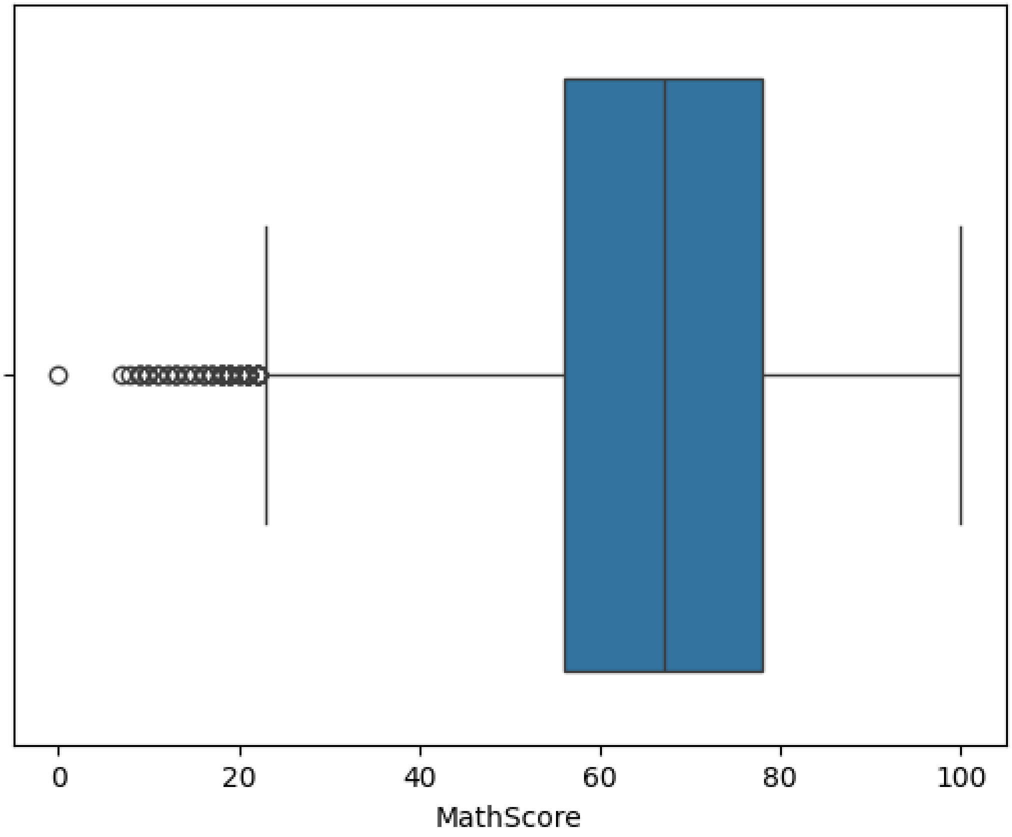
```
                     MathScore  ReadingScore  WritingScore
ParentMaritalStatus
divorced             66.691197     69.655011     68.799146
married              66.657326     69.389575     68.420981
single               66.165704     69.157250     68.174440
widowed              67.368866     69.651438     68.563452
```

```
In [43]: plt.figure(figsize= (5,5))
         sns.heatmap(gb1, annot = True)
         plt.title("Relationship btw parent's marital status & students score")
         plt.show()
```



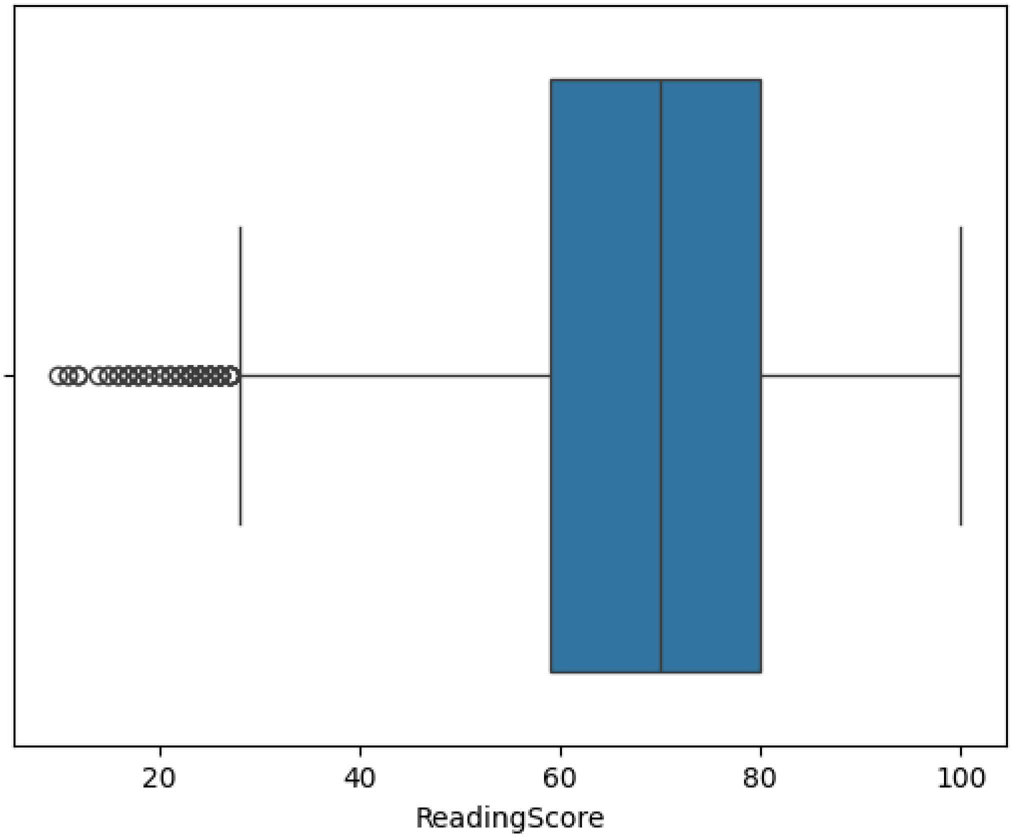
--From the above chart we have concluded that the parental maritalstatus do not have major impact or no impact there kids.

In [47]:
```
#deducting outliers(extreme values)
sns.boxplot(data = df, x = "MathScore")
plt.show()
```



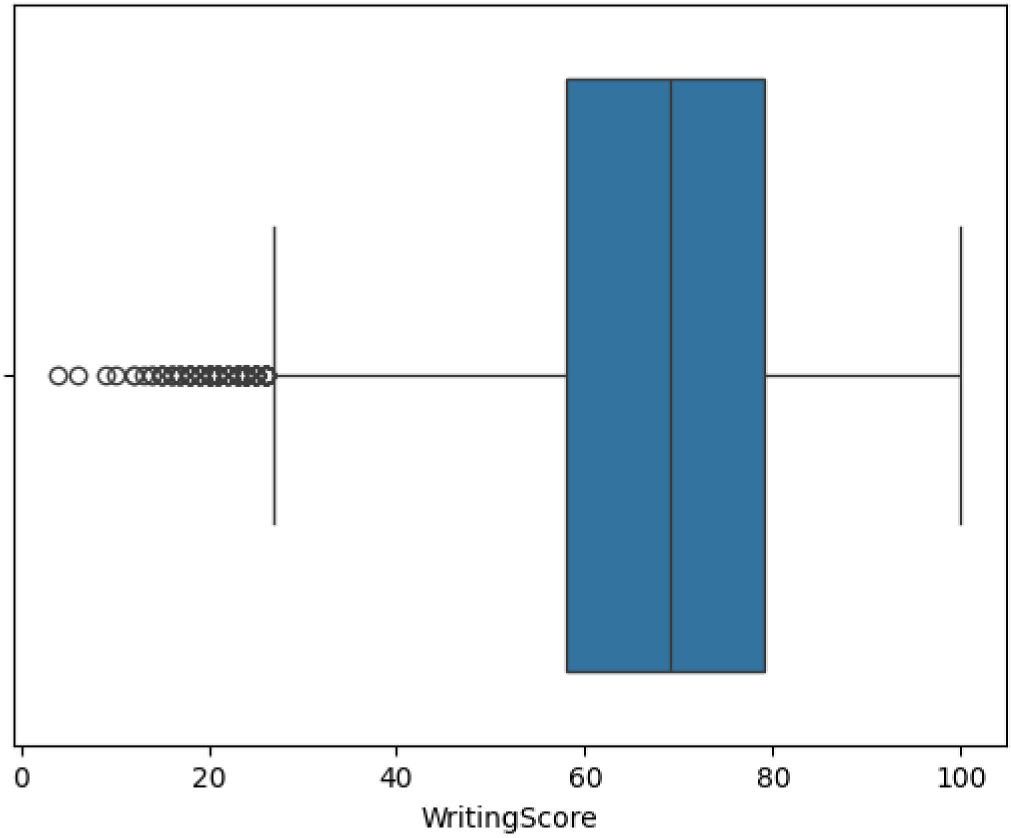
In [45]:

```
sns.boxplot(data = df, x = "ReadingScore")
plt.show()
```



In [46]:

```
sns.boxplot(data = df, x = "WritingScore")
plt.show()
```



```
In [48]: print(df["EthnicGroup"].unique())
```

```
[nan 'group C' 'group B' 'group A' 'group D' 'group E']
```

--Distribution of ethinic groups

```
In [51]: groupA = df.loc[(df['EthnicGroup'] == "group A")].count
         print(groupA)
```

```
<bound method DataFrame.count of        Gender EthnicGroup          ParentEduc  
LunchType   TestPrep  \
3        male     group A  associate's degree  free/reduced       none
13       male     group A        some college      standard  completed
14     female     group A     master's degree      standard       none
25       male     group A     master's degree  free/reduced       none
56       male     group A    some high school  free/reduced       none
...       ...         ...                 ...           ...        ...
30603    male     group A         high school      standard       none
30621  female     group A   bachelor's degree      standard       none
30622    male     group A  associate's degree  free/reduced  completed
30627  female     group A         high school      standard  completed
30634    male     group A  associate's degree  free/reduced  completed

      ParentMaritalStatus PracticeSport IsFirstChild  NrSiblings  \
3                 married         never           no         1.0
13                 single     sometimes          yes         1.0
14               divorced     sometimes          yes         2.0
25                married     regularly          yes         1.0
56                married     sometimes          yes         NaN
...                   ...           ...          ...         ...
30603              single     regularly           no         2.0
30621              single     regularly           no         2.0
30622            divorced     regularly           no         3.0
30627             married         never           no         NaN
30634                 NaN     sometimes           no         2.0

      TransportMeans WklyStudyHours  MathScore  ReadingScore  WritingScore
3                NaN           5-10         45            56            42
13           private           > 10         80            73            71
14           private            < 5         48            53            58
25        school_bus           5-10         75            76            76
56        school_bus           5-10         39            39            34
...              ...            ...        ...           ...           ...
30603     school_bus           5-10         71            63            65
30621        private           5-10         66            80            78
30622        private           5-10         53            53            53
30627     school_bus           > 10         58            77            82
30634     school_bus           5-10         65            60            60

[2219 rows x 14 columns]>
```

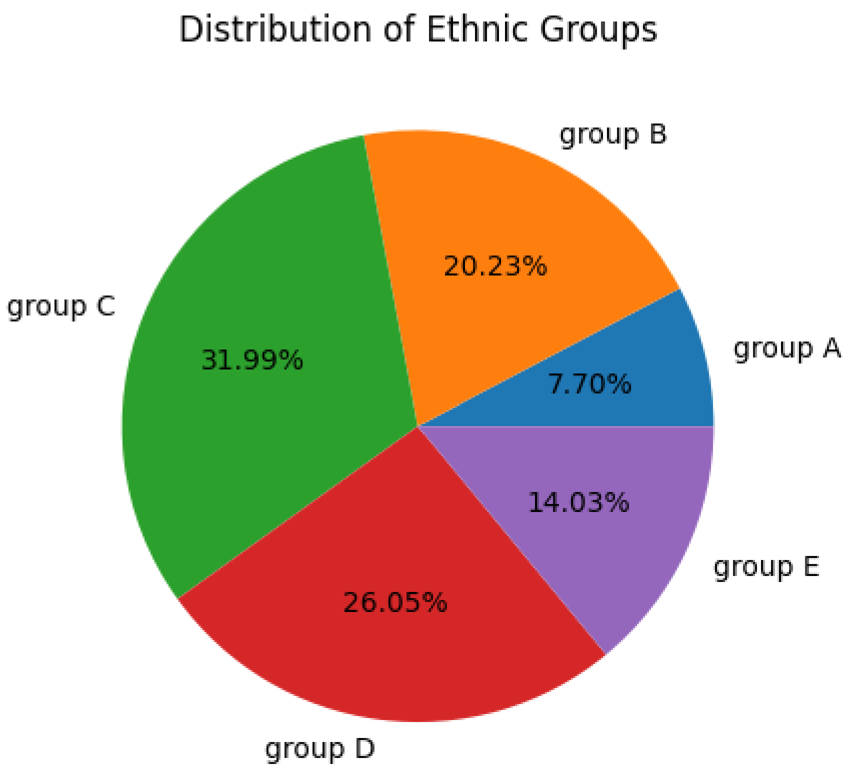
In [65]:

```python
groupA = df.loc[(df['EthnicGroup'] == "group A")].count()
groupB = df.loc[(df['EthnicGroup'] == "group B")].count()
groupC = df.loc[(df['EthnicGroup'] == "group C")].count()
groupD = df.loc[(df['EthnicGroup'] == "group D")].count()
groupE = df.loc[(df['EthnicGroup'] == "group E")].count()

l = ["group A", "group B", "group C", "group D", "group E"]
mlist = [groupA["EthnicGroup"], groupB["EthnicGroup"], groupC ["EthnicGroup"], g

print(mlist)
plt.pie(mlist, labels = l, autopct = "%1.2f%%")
plt.title("Distribution of Ethnic Groups")
plt.show()
```

```
[np.int64(2219), np.int64(5826), np.int64(9212), np.int64(7503), np.int64(4041)]
```

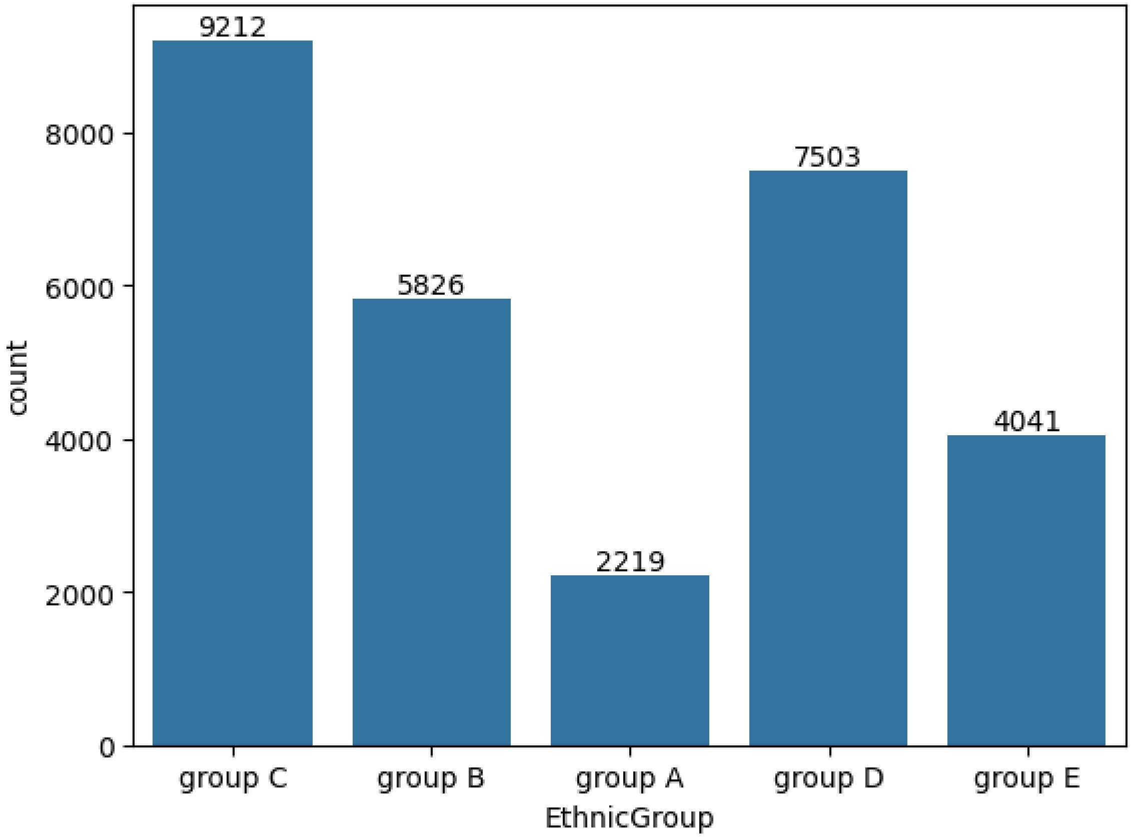


In [61]:

```
ax = sns.countplot(data= df, x = 'EthnicGroup')
ax.bar_label(ax.containers[0])
```

Out[61]:

```
[Text(0, 0, '9212'),
 Text(0, 0, '5826'),
 Text(0, 0, '2219'),
 Text(0, 0, '7503'),
 Text(0, 0, '4041')]
```



```
In [ ]:
```